NMR Study of Rpp30, an Archaeal RNase P Protein

A Senior Honors Thesis

Presented in Partial Fulfillment of the Requirements for
Graduation with distinction in Biochemistry in the undergraduate colleges of
of The Ohio State University

By

Christopher J. Bohlen

*****

The Ohio State University
June 2006

Project Advisor: Professor Mark Foster, Department of Biochemistry

ABSTRACT

RNase P, an essential enzyme, is an exceptionally interesting system for the study of protein-protein and protein-RNA interactions in catalysis. Composed of both RNA and protein subunits, the complex catalyzes the hydrolytic cleavage of the 5' leader sequence of precursor tRNA, a vital step in tRNA processing. While bacterial RNase P is a true ribozyme with a catalytic RNA subunit, it requires a single associated protein for function *in vivo*.

RNase P from archaeal and eukaryal organisms are also composed of requisite protein and RNA subunits. Coordinating at least nine proteins, eukaryotic RNase P is much more complex than the single-protein bacterial system, although both utilize a single RNA subunit. The archaeal enzyme is of intermediate complexity with at least four protein subunits. The functional roles of the RNA and protein subunits are reasonably well characterized in the bacterial system, but how these roles differ in organisms that use more protein subunits remains a mystery. Archaeal RNase P, then, is clearly a prime target for furthering our understanding of this enzyme in particular and of macromolecular cooperation for catalysis in general.

Pfu Rpp30 is the largest protein subunit of RNase P in the archaebacterium *Pyrococcus furiosus* (*Pfu*). Here, we show that Rpp30 has been recombinantly expressed and isolated by two different purification methods. NMR spectra of the isolated protein revealed dispersed signals characteristic of a well-folded protein, and optimization of spectral quality and subsequent three-dimensional experiments allowed ~90% of 206 assignable backbone amide resonances. These data were subsequently used to

characterize an interaction with Pop5, a protein binding partner of Rpp30. Widespread chemical shift perturbations observed for each protein's NMR spectrum suggest the formation of a tight complex, and mapping perturbations onto three-dimensional structures of Pop5 and Rpp30 indicated a previously unknown (and since corroborated) binding interface between the two molecules.

A detailed model of the Rpp30-Pop5 interface will be useful in designing future experiments to fully understand the functional contribution of RNase P subunit interactions. In addition, having high quality and assigned NMR spectra of Rpp30 opens the door for a detailed investigation of its involvement with other subunits using various NMR techniques.

ACKNOWLEDGEMENTS

As an undergraduate student, I would not have been able to make any of the progress here reported without constant nurturing from my labmates and advisors. All of the members of the lab have been unhesitant in providing me with the extensive guidance that my inexperience has demanded. I cannot thank them enough for the time they have invested in educating me. Just as importantly, the positive attitude and cohesiveness of the entire lab has been truly inspirational for the present research project as well as my long-term goals.

VITA

May 11, 1984……..…..Born Columbus, OH

2002-present…….Undergraduate Student

The Ohio State University

PUBLICATIONS

Wilson RC, Bohlen CJ, Foster MP, Bell CE. "Structure of Pfu Pop5, an archaeal RNase P protein." *Proc. Nat. Acad. Sci. USA.* (2006) **103** (4): 873-878.

FIELDS OF STUDY

Major Field: Biochemistry

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 RNase P:

RNase P is perhaps best known from the 1989 Nobel Prize in Chemistry[1]. That year, Sidney Altman was awarded for discovering that RNase P, a ribonucleoprotein complex, is a true RNA catalyst[2,3]. Namely, he showed that the RNA subunit of bacterial RNase P is catalytic on its own, demonstrating that the RNA (and not the protein) is the enzyme's catalytic component. Now, over twenty years after this initial discovery, the actual machinery of RNase P catalysis is still under investigation.

Despite the complexity of its mechanism, RNase P performs a straightforward biological function; the enzyme catalyzes the hydrolytic cleavage of precursor tRNA[4]. In order for freshly transcribed tRNA to become mature and functional, several post-transcriptional modifications must take place. RNase P performs one of these necessary modifications by removing a 5' leader sequence from precursor tRNA[4,5](**Fig 1.1**). Considering the variety among the 64 recognized tRNA sequences[5], RNase P achieves surprising accuracy in cleavage site selection. *In vivo,* this straightforward cleavage step is essential for proper tRNA functioning, and is thereby also vital for any living cell.

Accordingly, RNase P can be postulated to be necessary in all organisms, and it has been found in all domains of life[6]. Although several aspects have been conserved over evolutionary time, the enzyme has a particularly interesting developmental history.

In bacteria, RNase P consists of one large RNA subunit (~120kDa, referred to as 'PRNA') and a small protein cofactor (~15kDa). Altman's initial experiments and corroborating studies[2,7,8] indicate that bacterial PRNA is catalytically active *in vitro*

without protein. The protein is quite important, however, because it improves the enzyme's turnover rate [9,10,11], lowers the *in vitro* magnesium requirement [12], and is required for *in vivo* activity [13]. Presently, the protein subunit is hypothesized to be involved in some or all of the following: stabilizing the PRNA structure, recognizing the tRNA substrate, and promoting dimerization of the holoenzyme[14].

The RNase P complex in eukaryotes is surprisingly quite different from the bacterial enzyme despite its conserved function[15]. There remains a single large RNA subunit that shares sequence homology with bacterial PRNA. However, secondary structure mapping based on conserved sequence elements indicates that several structural components of bacterial PRNA are absent in its eukaryotic counterpart[16,17]**(Fig 1.2**), which may be the reason eukaryotic PRNA has not been demonstrated to be active *in vitro*. In addition to a truncated PRNA, the eukaryotic enzyme has at least nine associated protein cofactors (ranging in size from ~15-100 kDa)[18], none of which shares sequence homology with the bacterial protein[15]. Overall, the eukaryotic enzyme is not as well characterized as the bacterial enzyme due to this daunting complexity.

1.2 Archaeal RNase P:

Archaea, have an RNase P holoenzyme representing a hybrid of the eukaryotic and bacterial enzymes (**Fig 1.2**). There are two classes of PRNA in this domain. The "ancestral" Class A enzyme from most species of archaebacteria (including *Pyrococcus furiosus (Pfu)* and *Pyrococcus horikoshii (Pho)*) closely resembles the consensus bacterial PRNA, lacking only a few secondary structure elements[19]. Class A PRNA's from several archaeal organisms have been shown to be partially active in the absence of proteins *in vitro*, although only in solutions containing both monovalent and divalent ions

and having extreme (~4M) ionic strengths[20]. The second class of archaeal PRNA's, Class M, groups more truncated versions of the molecule that bear a closer resemblance to eukaryotic PRNA[19]. Accordingly, Class M PRNA requires protein subunits to perform catalysis *in vitro*[20]. Both classes can be reconstituted when some or all of four protein cofactors are included[21,22,(Gopalan unpublished data)]. A fifth protein has been identified in *Pho* that increases the optimum temperature of the reconstituted enzyme[23], but the protein is also involved with other ribonucleoprotein complexes and has not been implicated with other model organisms. In contrast, the other four proteins are homologous to four of the eukaryotic RNase P proteins, but they have no apparent sequence homology to the bacterial protein subunit[24]. In addition, the specific function of any of the four main proteins is presently unknown.

Thus, RNase P holds a number of evolutionary mysteries that invite further study. It is hypothesized that archaeal and bacterial domains have branched from a common ancestor that perhaps utilized a protein-free RNase P, accounting for the similarity in RNA subunits between the two domains despite the lack of homology in the protein components. The catalytic activity of homologous PRNA's isolated from these two domains of life compounded with their dissimilar protein complements lends credence to an "RNA World" model in which RNA catalytic machinery has been largely usurped by protein components over evolutionary time[25]. Eukaryotic RNase P represents further evolution on the archaeal branch, being composed of a number of homologous subunits and continuing the apparent progression from RNA to protein catalytic machinery. In RNase P, we have biomolecular "snapshots" of different stages in this progression, offering an exceptional model for studying ribozyme evolution.

Living near 100 ^{o}C thermal vents, *Pyrococcus furiosus* (*Pfu*) is a hyperthermophilic archaebacterium. Aside from having particularly thermostable biomolecules, *Pfu* RNase P is one of the best characterized RNase P holoenzymes from Archaea. The enzyme has been reconstituted *in vitro* as PRNA + four proteins (*Pfu* 1816, *Pfu* 1613, *Pfu* 1378 and *Pfu* 1914*,* which will be referred to by the names of their human homologues: Rpp29, Rpp21, Pop5 and Rpp30 respectively); partial subsets of the subunits have also shown catalytic activity (PRNA + Rpp21 + Rpp29 and PRNA + Pop5 + Rpp30)[(Gopalan, unpublished data)], which supports protein-protein interactions identified from yeast two-hybrid data observed for the closely related organisms *Pyrococcus horikoshii* (*Pho*)[26] and *Methanobacter thermoautotrophicus (Mth)*[27]. Gel retardation assays for *Pho* suggest some degree of association with PRNA for each of the four proteins[22]. Both protein-protein and protein-RNA interactions are further corroborated by yeast two and three-hybrid assays from homologues in the human system[28] (**Fig 1.3**).

Structural data from *Pfu* Rpp21[(Foster, unpublished data)], *Pfu* Pop5[29], and *Mth* Rpp29[21],are available and match structures of their homologues in *Archaeoglobus fulgidus (Afu)* and *Pho* [30,31,32,33,34]. Furthermore, a crystal structure of *Pho* Rpp30 has been solved[35]. The archaeal proteins all adopt canonical folds: Rpp29 has a beta-barrel fold, Rpp21 contains a zinc ribbon, Pop5 adopts the RRM (RNA recognition motif) fold, and Rpp30 is a TIM-barrel protein.

1.3 Understanding Subunit Interactions:

Ribonucleoprotein (RNP) complexes in general are extremely important in the cell, most evidently in the ribosome, but many details about the principles of protein-RNA interactions remain a mystery. A better understanding of RNase P's subunit

interactions will elucidate how protein components complement RNA roles in the enzyme[36]. Protein-RNA interactions are also involved in RNase P's amazing selectivity for tRNA-like substrates over other RNA's composed of the same four nucleotide types[37]. The enzyme's selectivity has inspired engineering efforts for targeted gene therapy and antiviral technologies based on the bacterial PRNA[38], and further information about protein roles in catalysis for any RNase P would surely be beneficial in this objective as well.

The intricate structures of RNase P and its substrate are fundamental to the overall reaction. In eukaryotes, RNase P has several RNA substrates besides tRNA, but all of them have a tRNA-like structural profile near the cleavage site[39]. In fact, deletion mutations in tRNA have been used to demonstrate that such structural elements (acceptor stem, T-stem, and T-stem loop) are necessary[40] and sufficient[41] for efficient RNA processing by bacterial RNase P. Much phylogenetic, footprinting, and crosslinking work has gone into modeling the bacterial RNase P complex and has succeeded in identifying a conserved PRNA active site and rough binding interfaces of the components[42,43,49]. Combined with structural data for the isolated components from several organisms[44,45,46,47,48], this research has culminated in a three-dimensional model of the holoenzyme[49](**Fig 1.4**).

Subunit interactions of RNase P from Archaea and Eukaryota are not so well-characterized. The eukaryotic enzyme has been partially reconstituted from cell extract[50], but a complete reconstitution using recombinantly expressed proteins has yet to be achieved. A putative interaction network has been derived for the eukaryal holoenzyme

from yeast two- and three-hybrid screens[28], and the only structural information about the complex comes from database mining[51] or homology modeling after archaeal proteins.

Study of archaeal RNase P is further along; the enzyme has been reconstituted with four or fewer proteins, tentative subunit interactions have been identified by yeast two- and three-hybrid screens, and the structures of all four proteins have been determined (as described in section 1.2). Still, a robust three-dimensional model of the archaeal holoenzyme is lacking and with it, a detailed model of the mechanism of the enzyme.

1.4 NMR Spectroscopy of Proteins[52,53]:

X-ray crystallography is by no means the only technique for characterizing macromolecular structures, and NMR spectroscopy of biological macromolecules is a robust and growing field as well. Although these techniques have almost nothing in common in terms of data collection, both can be used to obtain 3D structural models of macromolecules. Collection of NMR data, although generally more time-consuming than X-ray data, also provides information about the dynamics of the observed molecule in solution, which is potentially much more insightful than static properties observed in the artificially rigid confines of a crystal.

One of the most ubiquitous means of observing proteins with NMR is via the ^{15}N HSQC (Heteronuclear Single Quantum Correlation) spectrum. The HSQC pulse sequence first excites ^{1}H nuclei, then administers a series of pulses that transfers the excitation from the hydrogen to an adjacent ^{15}N nucleus. Varying the delay between a pair of ^{15}N pulses in this sequence allows the ^{1}H signals to be modulated according to the chemical shift of the attached ^{15}N, giving rise to the second dimension of the spectrum.

The reason this experiment is so fundamental to biomolecular NMR study is that it detects only protons covalently bonded to ^{15}N atoms. Although ^{15}N is found in very low natural abundance, a biomolecular sample can be isotopically enriched with ^{15}N or other NMR-active nuclei by growing recombinant cells in isotopically enriched media. For ^{15}N-enriched proteins, an HSQC spectrum gives information about every amide proton on the backbone as well as side-chain amide protons.

Using a variety of three-dimensional spectra with uniformly labeled ^{15}N- and ^{13}C-labeled protein, each of the amide resonances that appear on the HSQC spectrum can be assigned to its corresponding position in the protein's primary sequence (discussed in chapter 4). Obtaining "assignments" for the HSQC spectrum is the critical first step for translating NMR data into biomolecularly relevant information. In addition to being a major step toward structural characterization of a protein, an assigned HSQC allows the spectroscopist to take advantage of a number of techniques that describe the status of molecules in solution. For example, well-structured versus unstructured regions can be gauged based on signal intensity, and secondary structural features can be predicted based on database comparisons of the observed chemical shifts.

Particularly relevant for our purposes with RNase P, protein or nucleic acid atoms will experience a change in chemical environment upon interacting with other molecules (either from direct interaction or conformational changes), which translates into spectral changes. Thus, a binding event can be directly observed when a ligand or binding partner is added. If the changes are recorded for an assigned HSQC, the perturbations can be used to identify what residues are involved in the binding event, which, when combined with structural data, can be used to describe the location of binding sites. For example,

such experiments were conducted for *Pfu* Rpp29[21], which identified a strong Rpp29/PRNA interaction and located a putative binding site on the protein (**Fig 1.5**).

NMR characterization of each of the RNase P components will be useful for piecing together a three-dimensional model of the holoenzyme. Not only does the approach afford detailed information about biomolecular interactions, but it affords a powerful source of data by which future hypotheses (about, say, assembly pathways or involvement of particular residues) can be readily tested.

The power of NMR is not unlimited, however, and there are several substantial caveats to be considered. Perhaps most relevant to the present study is the limited range of molecular sizes from which well-resolved data can be obtained. First, a larger molecule will be composed of more atoms, which will give rise to more spectral peaks; as spectral resolution is limited, large numbers of peaks will eventually begin to overlap, becoming indistinguishable. This problem is confounded by the fact that larger molecules tumble more slowly in solution, which results in faster relaxation of spins and a decrease in NMR signal. Since the relaxation time is inversely proportional to the correlation time of a molecule (τ_c), the size limitation of NMR is roughly illustrated by the Stoke's equation, which gives τ_c for a sphere:

$$\tau_c = \frac{4\pi\eta_w r_H^3}{3k_B T}$$

(η_w is the viscosity of the solvent, r_H is the hydrated radius, k_B is Boltzmann's constant and T is the temperature[54]). As the size and effective radius of the molecule increases, so does the correlation time, which means it will take longer for the molecule to reorient, leading to a decay of the NMR signal (which is dependent upon randomized orientation).

NMR characterization of proteins and complexes has typically been limited to molecules of less than ~30 kDa. However, technical improvements are constantly advancing the field. Larger magnets, cryogenic probes, and TROSY (transverse relaxation-optimized spectroscopy) pulse sequences that are specialized for large molecules are making NMR characterization of larger molecules more tractable[55,56]. Study of hyperthermophile-derived archaeal RNase P can take advantage of high-temperature experiments (heat decreases τ_c, which is apparent from the Stoke's equation above), and so with selective labeling or fragmentation of the complex, NMR characterization is a challenging but feasible endeavor.

1.5 Specific Aims:

The long-term goal of the present study is to clarify the nature of archaeal RNase P subunit interactions using information about molecular structure and dynamics that is accessible by NMR spectroscopy. The structure-function relationships among interacting subunits will enable the construction of a 3D functional model and provide a launchpad for targeted manipulations and a precise understanding of the enzyme.

This dissertation reports progress made toward understanding archaeal RNase P through work on the *Pfu* Rpp30 subunit. Namely, completion of or significant progress on each of the following aims will be discussed in the following chapters.

Aim 1: Recombinantly express and purify the protein. The initial objective was to analyze whether well-folded Rpp30 could be harvested from *E. coli* in large enough quantities and purified to a suitable degree for further detailed characterization.

Aim 2: Establish and optimize a high-yield purification. As NMR was to be the primary tool for structure-function characterization of multiple interactions, a preparation protocol

yielding protein in very high quantity and purity was essential to counterbalance the costs of isotopic media and the reagent demands of NMR.

Aim 3: Characterize interactions between Rpp30 and other RNase P subunits. NMR spectroscopy of Rpp30 in isolation and in complex with other subunits can be used to describe changes that take place in portions of the molecule that engage with binding partners; this can be accomplished without completely solving the *Pfu* Rpp30 structure using a crystal structure of the closely related (68% identity, 82% similarity) *Pho* Rpp30 as a model (**Fig 1.6**).

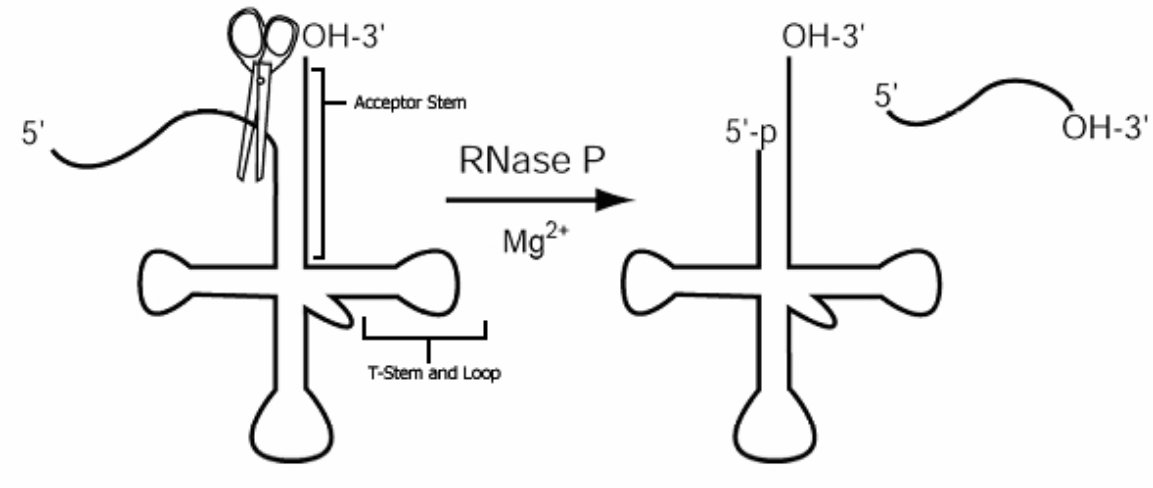

Figure 1.1: RNase P catalyzes tRNA maturation through cleavage of the 5' leader sequence. The enzyme recognizes the acceptor stem, T-stem, and T-loop elements of the substrate.

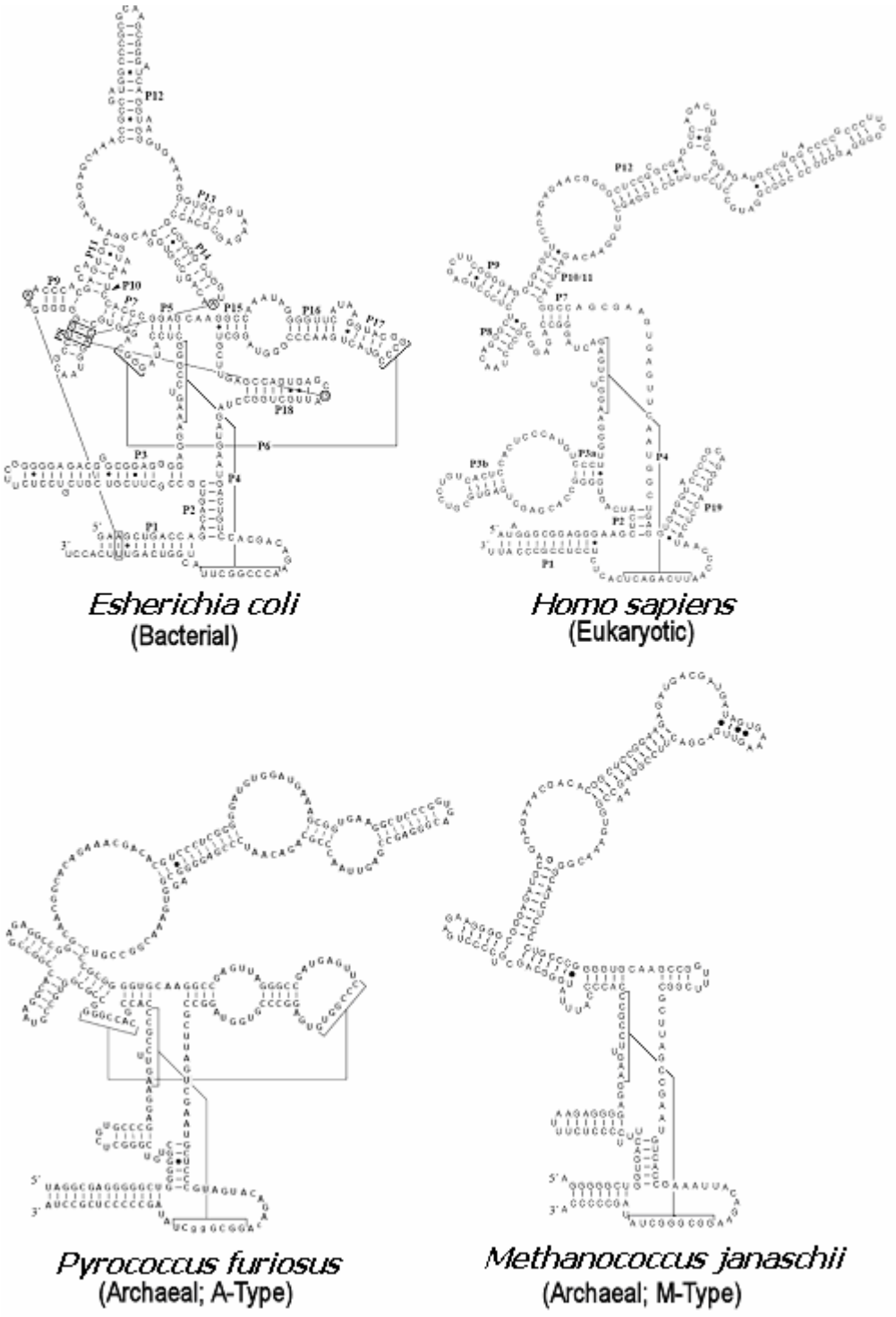

Figure 1.2: RNA subunits of RNase P from different organisms illustrate similarities and differences in predicted secondary structure. Major paired regions (P1-P19) are labeled on bacterial and eukaryotic RNA's. Archaeal RNA's are marked at the universally conserved P4 pairing region (adapted from[16]).

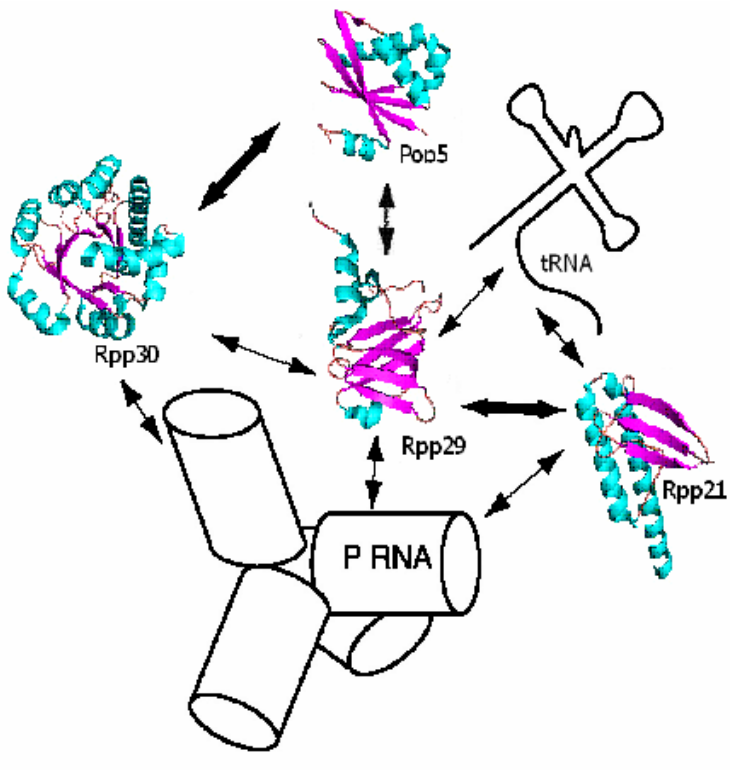

Figure 1.3: Interactions between archaeal RNase P proteins and PRNA subunits are starting to come to light. Three-dimensional structures of each of the four proteins are available. Those shown are from *Pho* except Pop5, which is from *Pfu*. Arrows indicate interactions identified from yeast two- and three-hybrid assays in human RNase P[28]. Other methods have indicated robust protein interactions in the archaeal system, identifying Rpp30-Pop5 and Rpp29-Rpp21 complexes (indicated by bold arrows). [26, 27, 33, 34]

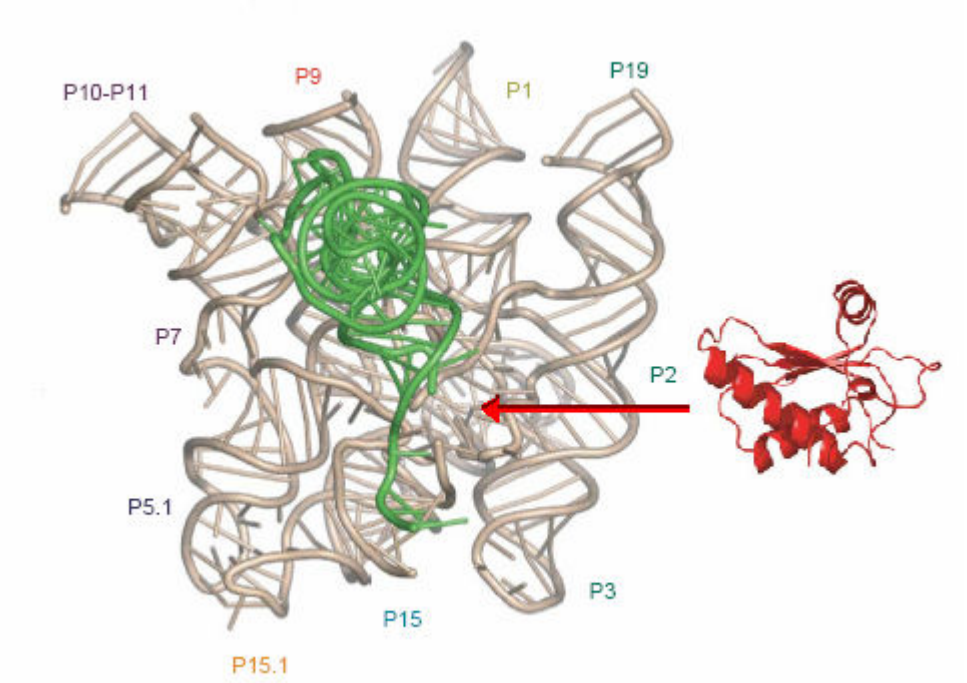

Figure 1.4: A recently proposed model of the bacterial holoenzyme illustrates the size ratios of the PRNA (gray), tRNA (green), and protein (red) subunits based on crystallographic structures of the isolated molecules[48]. The red arrow points to the protein binding site as identified by crosslinking and footprinting experiments. Similar studies have guided the positioning of the RNA in the model.

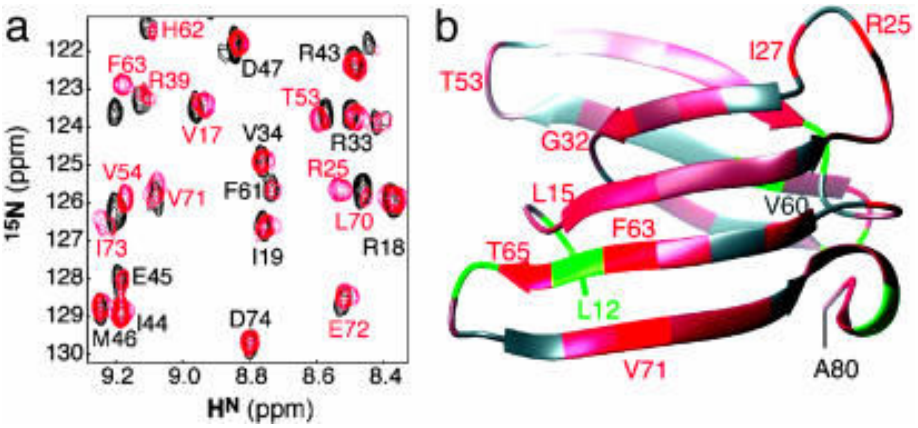

Figure 1.5: An example of using HSQC chemical shift perturbations (left) between free (black) and bound (red) states mapped onto an NMR structure of *Pfu* Rpp29 (right). The magnitude of the peak movement has been mapped as a gradient from gray to red (green residues could not be determined)[21].

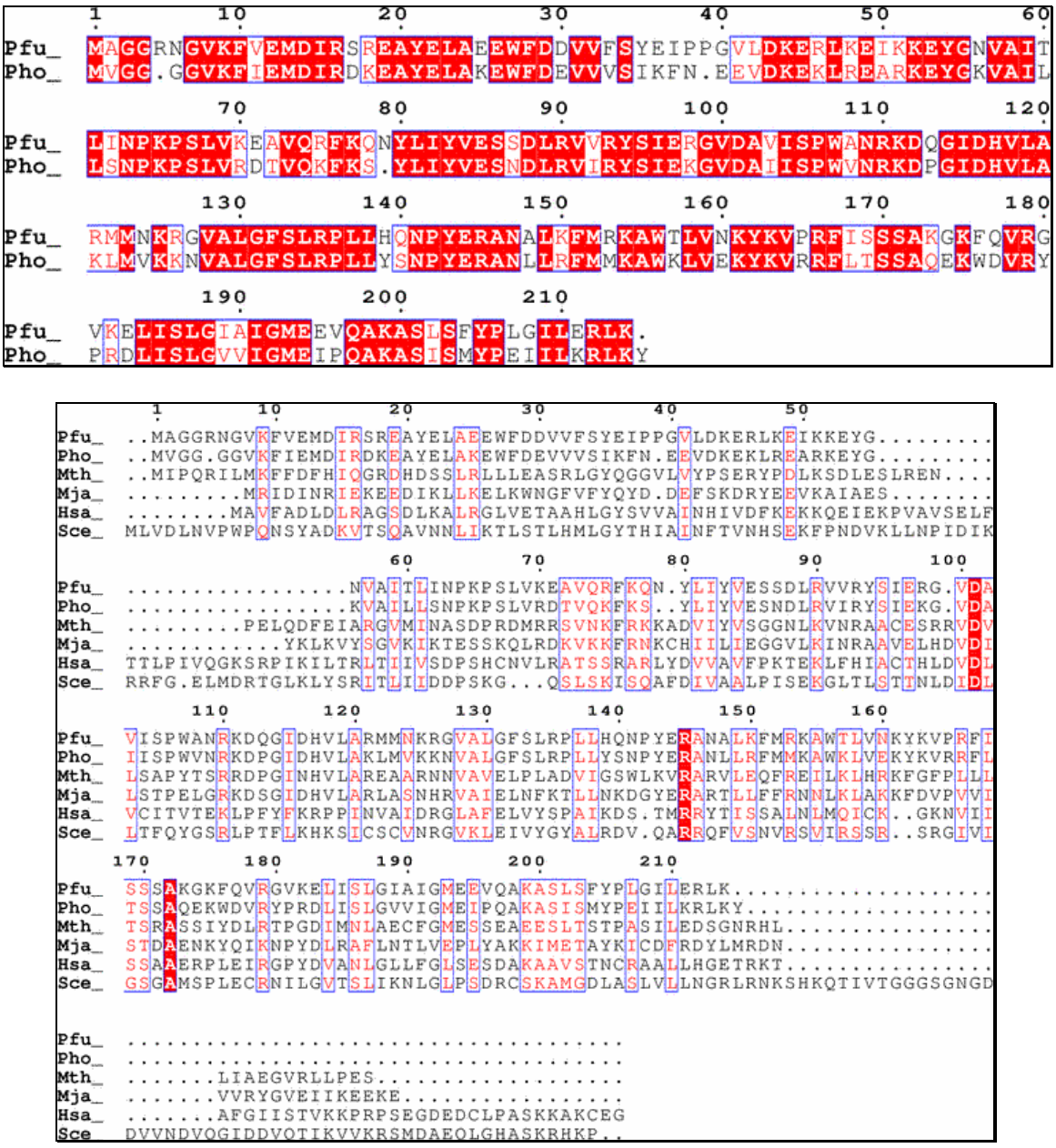

Figure 1.6: Sequence alignment between *Pfu* Rpp30 and *Pho* Rpp30 in the top panel shows close homology between the proteins. The bottom panel aligns Rpp30 homologues from four archaebacteria (*Pfu, Pho, Mth,* and *Mja)* and two eukaryotes (*Homo Sapiens(Hsa)* and *Saccaromyces Cerevisceae (Sce)*). Red-shaded residues are conserved and blue-framed residues have a global similarity score >0.7. Alignment was performed using CLUSTALW[67] and colored according to similarity by using a Risler scoring matrix[68] using the program ESPript[69].

CHAPTER 2

INITIAL CHARACTERIZATION OF *Pfu* Rpp30

2.1 Introduction:

Pyrococcus furiosus (Pfu) is one of the three archaebacteria from which RNase P activity has been reconstituted *in vitro*. Enzymatic activity has been observed in RNase P isolated from cultured *Pfu* cells; recombinantly expressed proteins combined with RNA transcribed *in vitro* have also been successfully reconstituted into functional RNase P[21]. Like other archaebacteria, this hyperthermophile has at least four RNase P proteins (Rpp29, Rpp21, Pop5, and Rpp30). In our laboratory, structures have been solved for *Mth* Rpp29[21] and *Pfu* Pop5[29], and structural characterization of *Pfu* Rpp29 is well underway, leaving only the Rpp30 subunit unsolved. Although there is substantial evidence for important interactions amongst the archaeal RNase P subunits, our understanding of their individual roles in the holoenzyme is incomplete.

The present project set out to address the functional and structural role(s) of the Rpp30 subunit. Although not absolutely essential for *in vitro* cleavage activity by the RNA subunit, only ~1% enzymatic activity has been observed in the absence of this critical component[22]. Detailed structural information will be important in refining our understanding Rpp30's role, but the first step in the characterization of a protein is its isolation, and NMR study demands particularly large amounts (~0.5 μmol) of protein. This objective seemed quite tractable due to some advantageous properties of Rpp30: 1) the protein is derived from a hyperthermophile and should therefore be stable in room-temperature purification and high-temprerature NMR experiments; 2) the protein has a very high isoelectric point (9.7) thereby facilitating purification; and 3) the protein has no

cysteine residues, eliminating the possibility of nonspecific inter- or intra-molecular disulfide formation. Although these factors did not result in facile purification, the initial steps of the project succeeded in overexpressing *Pfu* Rpp30 in *E. coli* and subsequently purifying it in its native form.

2.2 Materials and Methods:

2.2.1 Protein Expression

The *Pfu* Rpp30 gene (PF1914 from *Pyrococcus furiosus* DSM 3638) was first amplified from *Pfu* genomic DNA provided by Dr. Adams' laboratory (U. of Georgia) by using PCR with gene-specific primers. The gene was subsequently cloned into the bacterial expression vector pET33b by Walter Woznick (Dr. Gopalan's laboratory, OSU) with a standard restriction digestion and ligation protocol (NEB). *E. coli* BL21(DE3) cells were transformed with the plasmid by electroporation and grown at 37°C in 1L flasks of LB medium supplemented with 34 μg/liter chloramphenicol and 50 μg/liter kanamycin. The cells were induced by addition of 1 mM IPTG when the cells reached OD600 ≈ 0.6 (mid-log phase). After 4 hours of induction, cells were harvested by centrifugation (15 min, 5,000 × *g*, 4 °C), and stored at -20 °C. Thawed cells from 1 liter were suspended in 30 mL of lysis buffer (25 mM Tris, pH 7.3, 25 mM KCl) and sonication over ice was used to lyse the cells. Finally, soluble and insoluble components of the lysate were separated by centrifugation (15 min, 26,000 × *g*, 25 °C).

2.2.2 Protein Purification: Soluble Fraction

The soluble lysate was transferred to a fresh container where solid KCl and a dilute HCl solution were added to change the solution conditions to 25 mM Tris, pH 5.0, 1 M KCl. The solution was incubated at 25 ^{o}C for ~48 h. until precipitation of

contaminant proteins was complete. After separating the soluble and insoluble portions, the soluble fraction was added to an equal volume of water to dilute the KCl to 500mM. To precipitate nucleic acids, 10% w/w PEI stock was then added to a final concentration of 0.1% w/w, and the precipitate was removed by centrifugation. The soluble fraction was diluted to 375 mL (6.25-fold) with a new solution to a final buffer condition of: 40 mM KCl, 50 mM imidazole, 50 mM arginine, and 50 mM glutamate, 2 mM Tris, pH 7.0. This sample was passed through a 0.45 micron filter and loaded onto an equilibrated 5 mL HiTrap SP column (AP Biotech) and eluted with a salt gradient (0.04 to 2 M KCl). The purified protein product eluted in ~500 mM KCl and was confirmed to be ~90% pure Rpp30 by SDS PAGE visualized by Coomassie blue staining.

2.2.3 Protein Purification: Insoluble Fraction

The insoluble portion of the lysate was resuspended by sonication in 8 M urea, 25 mM Tris pH 10.5. Any fraction that was still insoluble was removed by filtration of the supernatant through a 0.45 micron filter. The sample was loaded onto a 1 mL HiTrap Q column (AP Biotech), and the flowthrough containing the protein was adjusted to a pH of 7.0 with dilute HCl. This solution was loaded onto an equilibrated 5 mL HiTrap SP column (AP Biotech). The protein was then buffer-exchanged and refolded by equilibrating the column with 80 mM KCl, 50 mM imidazole pH 7.0, 50 mM arginine, 50 mM glutamate. Finally, an elution gradient (0.08 to 2 M KCl) was run, eluting protein at ~500 mM KCl, which was confirmed to be ~95% pure Rpp30 by SDS PAGE visualized by Coomassie blue staining.

2.2.4 NMR Data Collection

Samples were obtained using the above purification protocols except that

transformed cells were grown in M9 minimal media supplemented with 1% Eagle Basal Vitamin Mix (Life Technologies, Gaithersburg, MD)[57], containing 1 g/liter ^{15}N ammonium chloride as the sole nitrogen source. The protein obtained via the soluble and insoluble protocols were separately concentrated to 500 μL and supplemented with D_2O to 5%. 2D ^{15}N HSQC spectra were recorded on a Bruker DMX-600 spectrometer at 28-55°C. Data were processed with NMRPipe[58] and NMRView[59].

2.3 Results and Discussion:

2.3.1 Sample Preparation

The Rpp30 gene was readily overexpressed in bacteria. Substantial protein production was observed, but the majority (~60%) of the protein formed inclusion bodies and was found in the insoluble fraction after lysis **(Fig 2.1)**. A high-yield preparation was desired, and so a purification protocol was devised for both the soluble and insoluble fractions of the lysate.

Initial attempts to purify the sample by heating were unsuccessful; the denatured contaminant proteins aggregated with native Rpp30, precipitating virtually all of the solution's proteins **(Fig 2.2)**. In addition, initial attempts to purify the protein by cation exchange were also unsuccessful. Although some pure protein could be eluted from the column, most of the protein did not bind to the resin **(Fig 2.2)**, despite the high PI of the protein and low salt concentration of the buffer. Because Rpp30 was already suspected to be involved in protein-RNA interactions[22,28], we hypothesized that the charged sites of the protein were occupied by other macromolecules or fragments, preventing binding to the cation exchange resin. Thus, we needed to remove a substantial portion of

contaminants (especially nucleic acid contaminants) before cation exchange purification could be achieved.

For the soluble fraction of the lysate, this was achieved using selective precipitation. Vapor diffusion assays[60] of small amounts of purified protein indicated that Rpp30 was particularly soluble at high salt concentrations. Thus, various conditions were screened and it was discovered that a low pH, 1 M salt combination precipitated most of the contaminant proteins over ~48 h. while minimizing Rpp30 precipitation. The partially purified sample was then subjected to PEI precipitation to remove nucleic acids. Again, the salt concentration was crucial; too much salt prevented PEI binding to nucleic acids, and too little salt caused the Rpp30 to precipitate along with the nucleic acids. Finally, the mostly-pure sample was diluted further into a low salt solution and loaded onto an SP column. As hoped, binding was no longer a problem, and the protein eluted midway through the salt gradient.

The insoluble fraction of the lysate could not be subjected to the same treatment as the soluble fraction because the urea required to resuspend the fraction prevented selective precipitation. However, since the protein was already denatured, the pH was increased to 10.5 to deprotonate as many amino acids as possible and disrupt interactions with the negatively charged nucleic-acids. Then, the solution was passed through an anion exchange column. The still negatively-charged nucleic acid fragments as well as the majority of the now negatively-charged contaminant proteins bound to the resin, while most of the Rpp30 was retained in the flowthrough. When the pH of the flowthrough was dropped to 7.0, the protein bound to cation exchange resin, although again only when virtually no salt was present. After binding Rpp30 to the resin of an SP

column, the protein could be refolded by washing non-denaturing buffer over the column, and subsequent elution yielded pure Rpp30.

In summary, a purification protocol was established that yielded moderately pure protein (>90% pure, as estimated by SDS PAGE) in fairly large quantities (~6 mg Rpp30 per liter of cell culture). Also, a collection of results strongly suggests that Rpp30 is associating with nucleic acids: the precipitation of Rpp30 with the addition of PEI in low salt but not in high salt, the inclusion of large amounts of nucleic acids in the insoluble fraction of the lysate, and the inability of Rpp30 to bind a cation exchange resin despite its high PI. Although association with nucleic acid may well be nonspecific given the high PI of the protein, the above results suggest that Rpp30 may readily bind RNA, as has been suggested before[22,28].

2.3.2 Preliminary NMR Data

Once a substantial quantity of protein had been purified, the next question was whether the product would be suitable for the desired NMR experiments. Before mixing the samples obtained from each purification protocol, we needed to determine whether the protein from the insoluble preparation refolded properly into the same conformation as the native protein. A rather low concentration sample of ^{15}N Rpp30 purified from each of the established protocols was analyzed using an HSQC pulse sequence; a ~150 μM sample from the insoluble preparation and ~250 μM sample from the soluble preparation were used.

An initial temperature screen was conducted for the insoluble preparation sample. Varying the temperature from 25-55 ^{o}C in 10 degree intervals showed that spectral quality improved with increasing temperature (**Fig 2.3)**; temperatures above 55 ^{o}C were

not used because higher temperatures would be damaging to the cryogenic probe that was planned for future experiments. Although spectral improvement may have been caused in part by decreased relaxation times, the improvement of signal was very large, suggesting that the hyperthermophile-derived protein was better structured at higher temperatures. Hanging-drop screens of buffer conditions suggested that little improvement in signal would be achieved by changing buffer components because the protein was maximally soluble at the high-salt, pH 7 condition already observed.

Using the temperature-optimized condition, spectra were compared for the protein samples derived from the soluble and insoluble preparations. The spectra were very similar, indicating that the refolded protein had the same conformation as the protein that had not been denatured **(Fig 2.4)**. Both spectra were well-dispersed. A wide range of ^{1}H chemical shifts spanning the 6.5-10.5 ppm indicated that the protein had assumed a well-defined tertiary structure. An unstructured protein, in contrast, would give a roughly homogenous signal spanning from ~7-8 ppm in the hydrogen dimension and resembling an unstructured string of amino acids.

Unfortunately, the overall clarity of the spectra was less than ideal. Only 158 peaks (including those from sidechains) were counted, which is substantially less than expected from the 214 backbone residues. There were large regions of overlap in the middle of both spectra, and the overlap was worse in the higher-concentration spectrum from the soluble preparation. Although this might have been caused by Rpp30 aggregation, it was also the case that the soluble preparation protein was slightly less pure than that from the insoluble preparation. Still, observation of the characteristic spectrum of a well-folded protein was an important result.

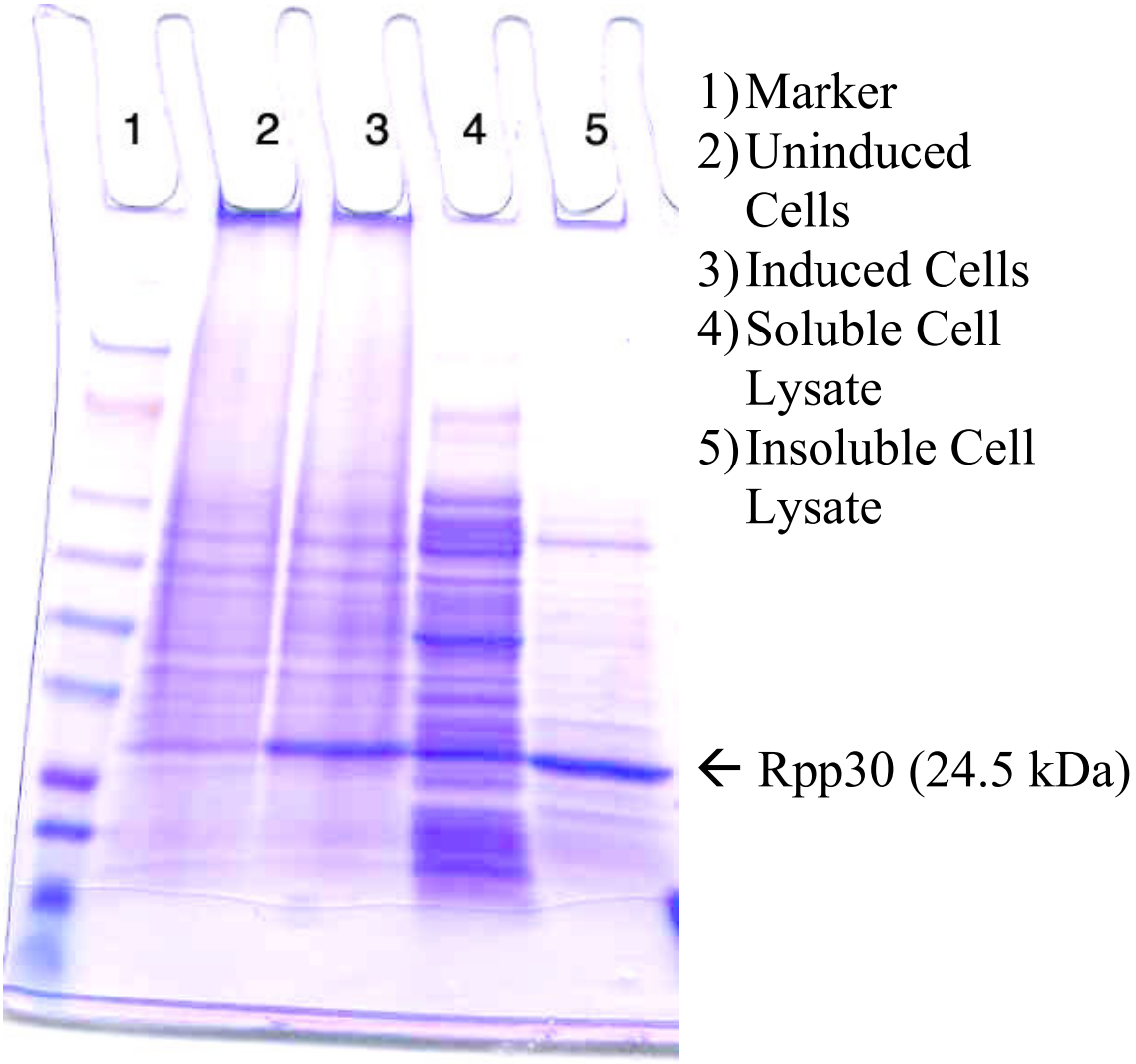

Figure 2.1: Induction and solubility of wild-type Rpp30 is illustrated by Coomassie-stained SDS-PAGE. Rpp30 can be overexpressed in *E. coli* (lanes 2 and 3), but is mostly present in the insoluble fraction of the lysate (lane 5).

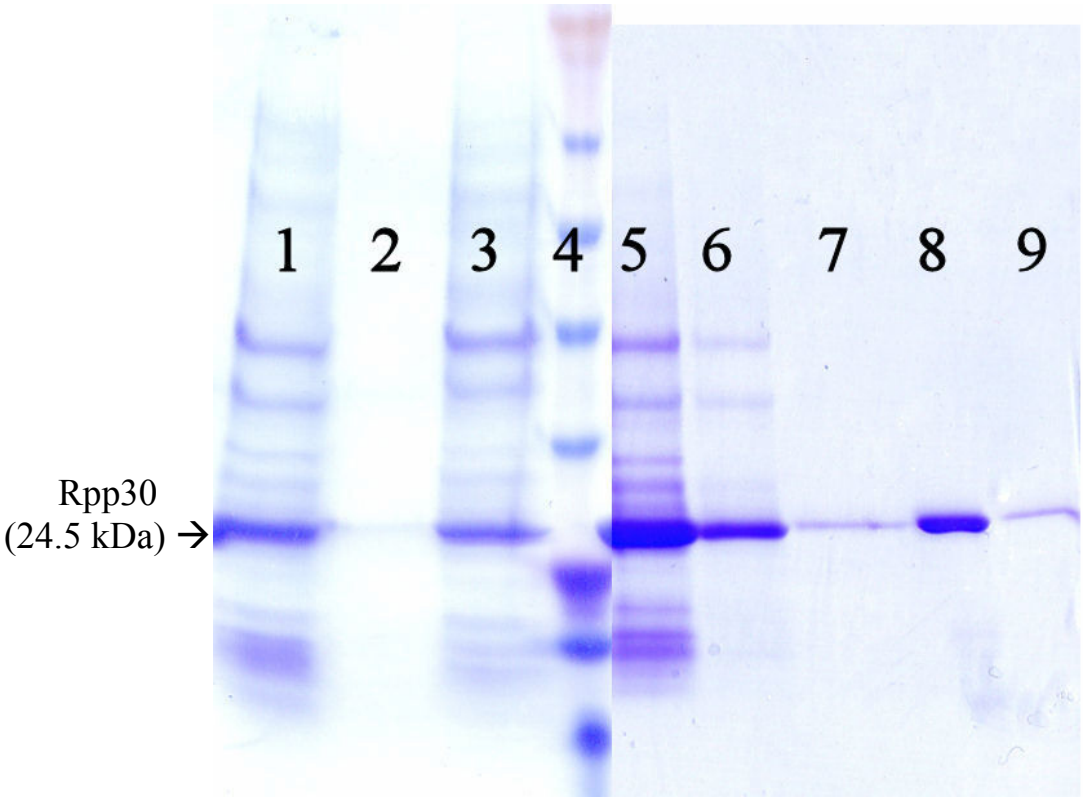

Figure 2.2: Purification of wild-type Rpp30 is described by SDS-PAGE. Lanes 1-3 show the inefficacy of heat-purification. Lane 1 is the soluble lysate, lane 2 is the soluble lysate after heating at 55^{o}C for 15 min, and lane 3 is the protein that precipitated during the heating process, which includes almost the entire protein content of the sample. Lanes 5-9 illustrate the pros and cons of SP cation exchange purification. Although some protein is collected in elution fractions (Lanes 7-9, concentrated), most Rpp30 loaded from the soluble lysate (Lane 5) flows through the column (Lane 6).

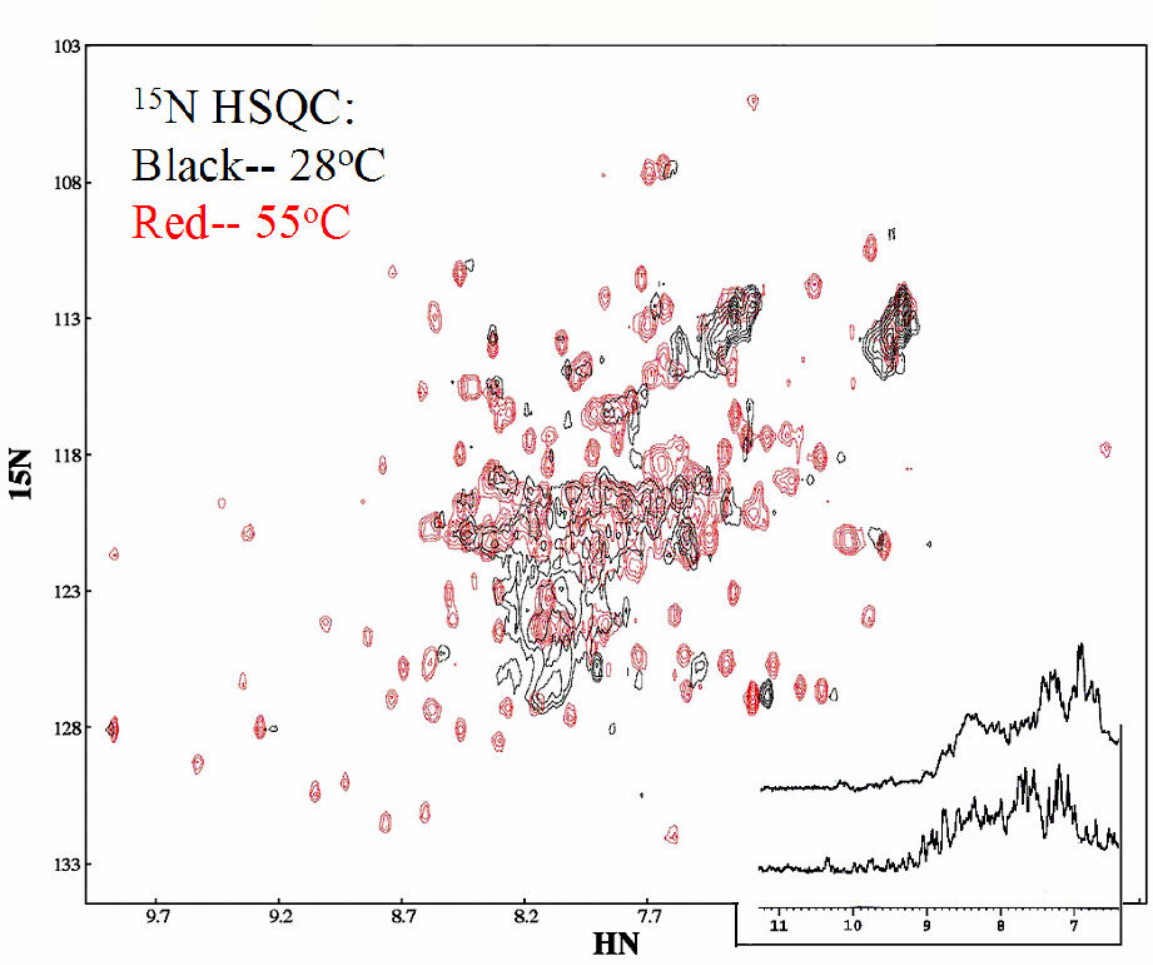

Figure 2.3: Spectral quality improves drastically with increased temperature. Improved dispersion and linewidth are evident in the amide region of Rpp30's ^{1}H spectrum (lower right inset) when collected at 28 ^{o}C (above) and at 55 ^{o}C (below). The improvement is even more pronounced in when ^{15}N HSQC spectra from the two temperatures (black spectrum collected at 28 ^{o}C, red spectrum at 55 ^{o}C) are overlaid. Spectra were collected in ~500 mM KCl, 50 mM imidazole pH 7.0, 50 mM arginine, 50 mM glutamate.

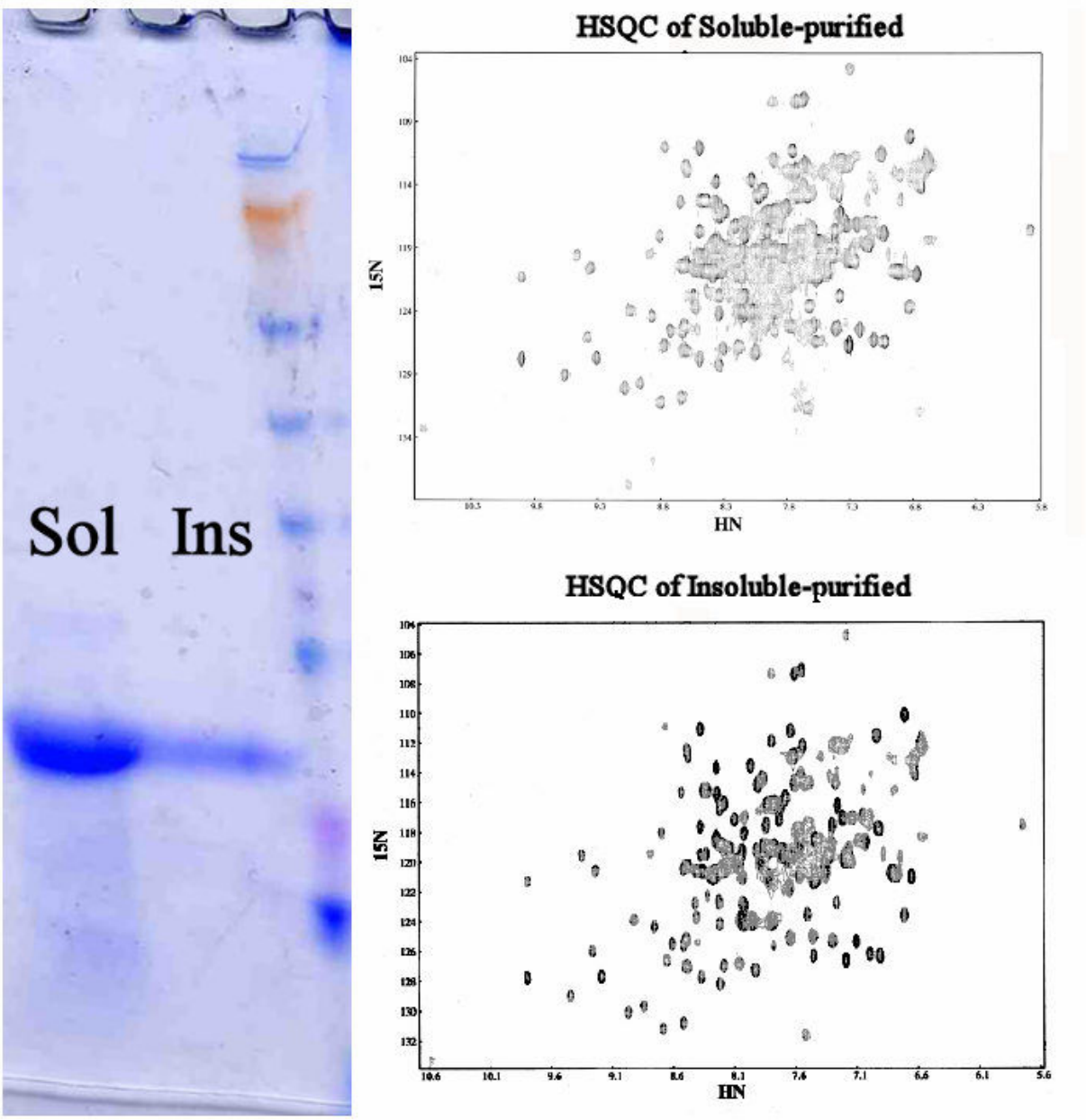

Figure 2.4: Sample homogeneity obtained purifying from the soluble (Sol) versus the insoluble (Ins) fractions of the cell lysate is visible by SDS-PAGE. A comparison of HSQC spectra shows overall similarity of fold, but the larger proportion of impurities in the soluble sample diminishes spectral quality (consider the degree of overlap in the center of the spectra) despite the higher concentration of the soluble-purified sample (250 μM vs. 150 μM).

CHAPTER 3

AFFINITY PURIFICATION OF AN MBP-Rpp30 FUSION PROTEIN

3.1 Introduction:

Pfu Rpp30 is a 24.5 kDa protein, which is on the larger end of the molecular-weight range for proteins amenable to structural characterization by NMR. As such, the demands for well-resolved spectra are particularly high. Despite the effort invested in achieving a moderately high-yield purification of the wild-type protein, the resulting HSQC spectra were still far from ideal; fewer than the expected number of peaks were observed, largely due to the obfuscating overlap in the central regions of the spectra. Because sample homogeneity is so important for NMR spectra, the small amount of impurities that could be detected in the sample were suspected to be diminishing the spectral quality. To improve the purity and yield of the protein, we therefore turned to a MBP (maltose-binding-protein)-fusion strategy to take advantage of the selectivity of affinity chromatography and the solubility-enhancing properties of MBP.

The pHMT plasmid (**Fig 3.1**) was employed to create a fusion of Rpp30 that could be purified by affinity chromatography. Three important features of the plasmid were advantageous. Directly upstream of the multiple cloning sites, the plasmid encodes the native *E. coli* maltose-binding-protein. Upstream from the MBP sequence there is a His6 tag, which is preceded by a start codon and promoter region. Between the His-tagged MBP and the multiple cloning sites, there is a region coding for the amino acid sequence that is recognized by the extremely specific tobacco etch virus (TEV) protease.

Rpp30 was cloned into pHMT to take advantage of all three features. Fusion to MBP, which is a highly soluble protein, has been suggested to aid the solubility of

overexpressed proteins[61]. The N-terminal His-tag offers a means of IMAC (immobilized metal affinity chromatography) purification, and after purification, both the MBP and the His-tag could be removed from Rpp30 using TEV Protease. MBP can be removed using amylose affinity chromatography as well. Furthermore, if $(His)_6$-tagged TEV Protease were used, both the cleaved MBP and the protease could be separated from Rpp30 from an additional IMAC step. Finally, the MBP-fusion form of the protein had potential for use in future pulldown experiments using insoluble amylose beads. We here show that this fusion form of the protein yielded a sample of improved purity and high spectral quality.

3.2 Materials and Methods:

3.2.1 pHMT Cloning:

The *Pfu* Rpp30 gene was subcloned from pET33b into the *Eco*RI and *Bam*HI sites of the pHMT vector[61] for expression as a fusion to His-6-tagged maltose binding protein (MBP); the construct is designed to allow removal of the fusion protein by specific cleavage with tobacco etch virus (TEV) protease. Subcloning was accomplished by PCR-based amplification of *Pfu* Rpp30/pET33b using the following oligonucleotide primers: 5'-GGAGA*GAATTC*ATGGCTGGTGGTCGTAATGG-3' and 5'-GACGG*GGATCC*TTATCATTTAAGACGTTCCAGAATACC-3', incorporating *Eco*RI and *Bam*HI sites at the 5' and 3' end of the gene, respectively (italicized); the start and stop codons are underlined. The identity of the subclone was confirmed by DNA sequencing (OSU PMGF Facility).

3.2.2 TEV Protease Purification:

The TEV protease used to remove the MBP tag was grown and isolated in-house.

The pRK793 plasmid, which encodes a mutant form of TEV protease that is resistant to self-cleavage, was obtained from Dr. David Waugh (NCI). A protocol also provided by Dr. Waugh[62] was used to guide purification of the protease.

BL21(DES)-RIL cells containing pRK793 were grown at 37 °C in 1 L flasks of LB-media containing 100 μg/ml carbenicillin and 30 μg/ml chloramphenicol. Cells were induced at $OD_{600} \approx 0.6$ with 0.5 mM IPTG and temperature was shifted to 30 ^{o}C for 4 h. Cells were centrifuged, then 1/3 of the pellet was resuspended in 10 ml of 50 mM PO_4 (pH 8.0), 100 mM NaCl, 10% glycerol, and 25 mM imidazole and lysed by sonication. PEI (10% w/w stock, pH 8.0) was added to the lysate before centrifugation. The supernatant was passed through a 0.45 micron filter then loaded onto an equilibrated 1 mL HiTrap Chelating column (AP Biotech) charged with $NiSO_4$. Contaminants were removed using a 0-200 mM imidazole gradient, then pure TEV protease was eluted using 500 mM imidazole solution. The product was diluted and DTT and EDTA were added to a final solution containing ~1 mg/mL TEV protease (estimated using Bradford reagent), 25 mM PO_4 (pH 8.0), 200 mM NaCl, 2 mM EDTA, 5 mM DTT, 10% glycerol, 50 mM imidazole. The final product was stored at -20 ^{o}C until needed.

3.2.3 Rpp30 Expression and Purification:

Escherichia coli BL21(DE3) cells were co-transformed with *Pfu* Rpp30/pHMT and pARG-U plasmids[63] and grown at 37°C in LB broth supplemented with 34 μg/liter chloramphenicol, 50 μg/liter carbenicillin. The cells were induced by addition of 0.5 mM IPTG when the cells reached $OD_{600} \approx 0.6$ (mid-log phase), at which point the culture was shifted to 30 °C. Cells were harvested after 20 h by centrifugation (15 min, 5,000 × *g*, 4 °C), and stored at -20 °C. Thawed cells from 1 liter were suspended in 30 ml of lysis

buffer (40 mM imidazole, pH 7.6/500 mM KCl), lysed on ice by sonication, mixed with 0.1% (wt/vol) PEI, and clarified by centrifugation (15 min, 26,000 × *g*, 25°C). The soluble lysate was passed through a 0.45 μm filter and loaded onto an equilibrated 5 ml HisTrap column (AP Biotech), then washed with 50 ml of lysis buffer containing 10% glycerol and 25 ml lysis buffer without glycerol. The MBP-fused protein was then eluted with (500 mM imidazole pH7.6/500 mM KCl). Fractions containing MBP-fused Rpp30 were pooled for dialysis against 30 mM imidazole pH 7.6/250 mM KCl. After 16 h of dialysis, the protein concentration was estimated from its denatured extinction coefficient at 280 nm, 28,590 $M^{-1} \cdot cm^{-1}$ (http://us.expasy.org/tools/protparam.html), and the solution was diluted with dialysis buffer to 80 μM protein. The fusion protein was then cleaved by incubation for 1.5 h at 30 °C with a 5% molar equivalent of His-tagged S219V-TEV protease[64]. After cleavage, KCl was added to 500 mM and Rpp30 was purified from the His-tagged MBP and TEV protease by passing the solution through a 5-ml HiTrap Chelating column (AP Biotech) charged with $NiSO_4$. The flow-through was diluted twofold with water and further purified on a 1 ml HiTrap SP column (AP Biotech) eluted with a linear salt gradient (0.25-2 M KCl). The final product was confirmed to be >95% pure by SDS PAGE visualized by Coomassie blue staining and ESI mass spectrometry (provided by OSU CCIC facility; 24827.9 Da expected for GEF-Rpp30 (GEF is residual from TEV cleavage site and *Bam*HI cloning site used for the fusion construct); 24,828 Da observed) before dialysis against NMR buffer (20 mM sodium acetate, pH 5.0/20 mM KCl). Typical yields of pure protein were around 13 mg per liter of rich media (LB).

3.2.4 NMR Data Collection

Samples were obtained using the above purification protocols except that transformed cells were grown in M9 minimal media supplemented with 1% Eagle Basal Vitamin Mix (Life Technologies, Gaithersburg, MD)[57], containing 1 g/liter ^{15}N ammonium chloride as the sole nitrogen source, and D_2O was added to the final sample to 5%. 2D ^{15}N HSQC and TROSY spectra were collected on a Bruker DMX-800 spectrometer with a cryogenic probe at 55 ^{o}C. Data were processed with NMRPipe[58] and NMRView[59].

3.3 Results and Discussion:

Creation of the MBP-Rpp30 fusion mutant allowed for a greatly improved purification (**Fig 3.2**). The protocol was simplified, the yield was roughly doubled, and there was no significant contamination observed by Coomassie-stained SDS PAGE or ESI mass spectrometry (**Fig 3.3**). Mass spectrometry also indicated that the desired protein product had been obtained (24827.9 Da expected for GEF-Rpp30; 24,828 Da observed). Solubility was also greatly improved in contrast to samples obtained from the wild-type purifications; before, Rpp30's solubility limit was very low (~150 μM) in low salt conditions, and only ~500 μM in 0.5 M KCl, but protein obtained from the affinity purification protocol could be concentrated to at least 1.5 mM in only 10 mM sodium acetate pH 5.0. Buffer screenings on this sample suggested that solubility was maximal at pH 5.0 in acetate buffer with little dependence on salt concentration.

The improvement of purity was also manifest in NMR spectra, where a 0.65 mM Rpp30 sample in low-salt gave very high-quality spectra from relatively short experiments (**Fig 3.4**). TROSY experiments were also collected, but the signal-to-noise

was slightly lower than the HSQC experiment. However, the TROSY does not detect sidechain nitrogens very well, and so an overlay of the spectra was used to identify 20 sidechain peaks (**Fig 3.5**). Subtracted from the 220 peaks that could be counted in the HSQC, this left 200 peaks that could be identified from a 2D spectrum, a large portion of the 214 expected resonances from backbone amides.

Overall, the quality of NMR data strongly supported further characterization of the protein. Not only could we be confident in obtaining useful 3D NMR data with relatively few scans, but the high purity of the sample suggested that experiments probing Rpp30 interactions would yield interpretable results.

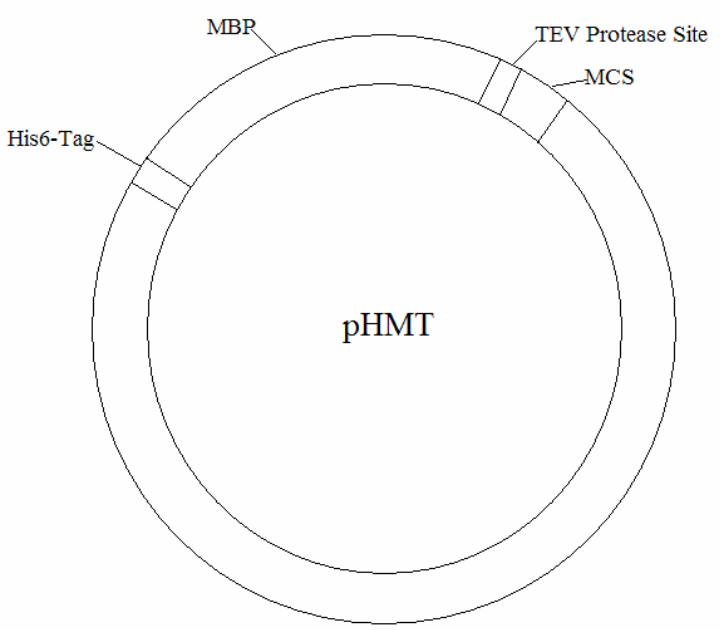

Figure 3.1: A plasmid map for pHMT shows the major features, including a His-tag, an MBP-encoding region, a TEV-protease recognition site, and a multiple cloning site.

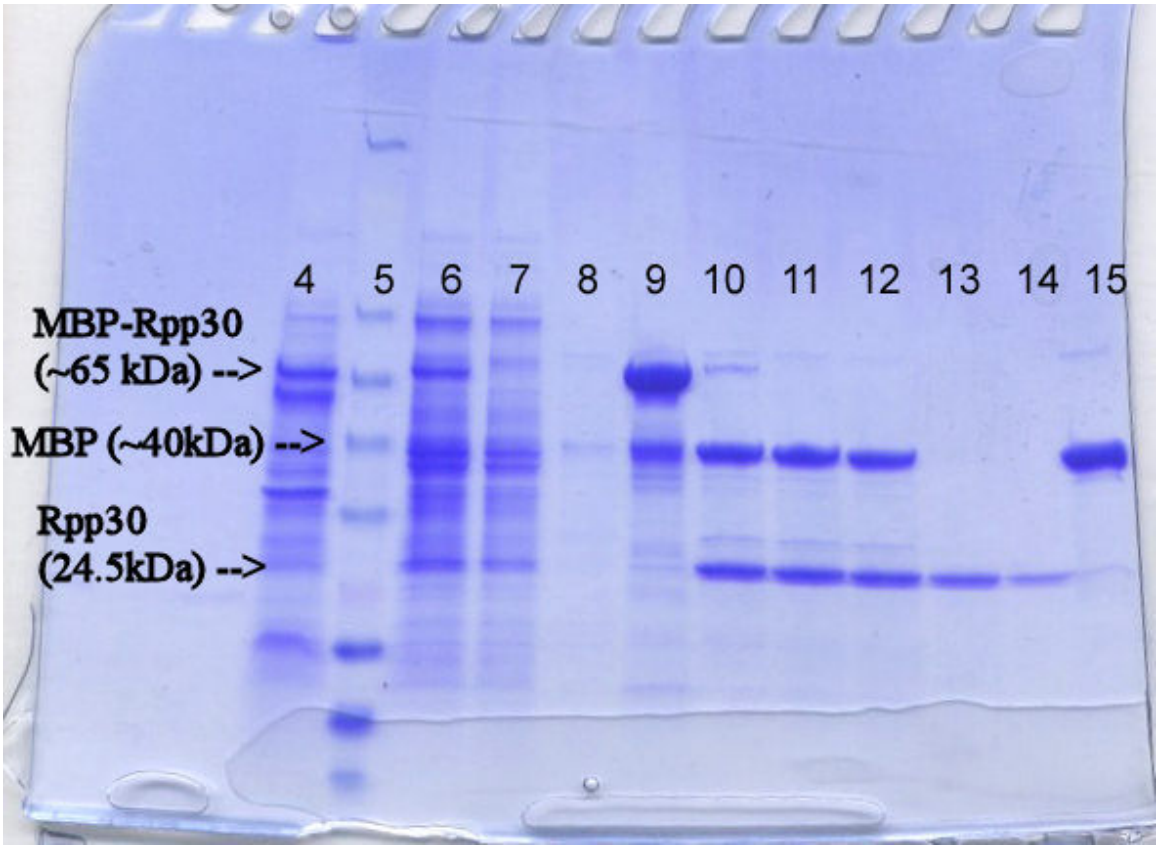

Figure 3 2: Purification of the MBP-Rpp30 construct was optimized. In this SDS-PAGE gel, lanes 4 and 6 compare the insoluble and soluble lysates respectively. Lanes 7 and 8 represent protein that flows through the first nickel column of the protocol, and lane 9 is the elution product. Lanes 10-12 are time points from TEV-protease digestion of the sample from lane 9, which were subsequently loaded onto a nickel column. Lanes 13 and 14 are the flowthrough and early elution fractions containing pure Rpp30, and lane 15 is a later elution fraction containing the cleaved MBP tag.

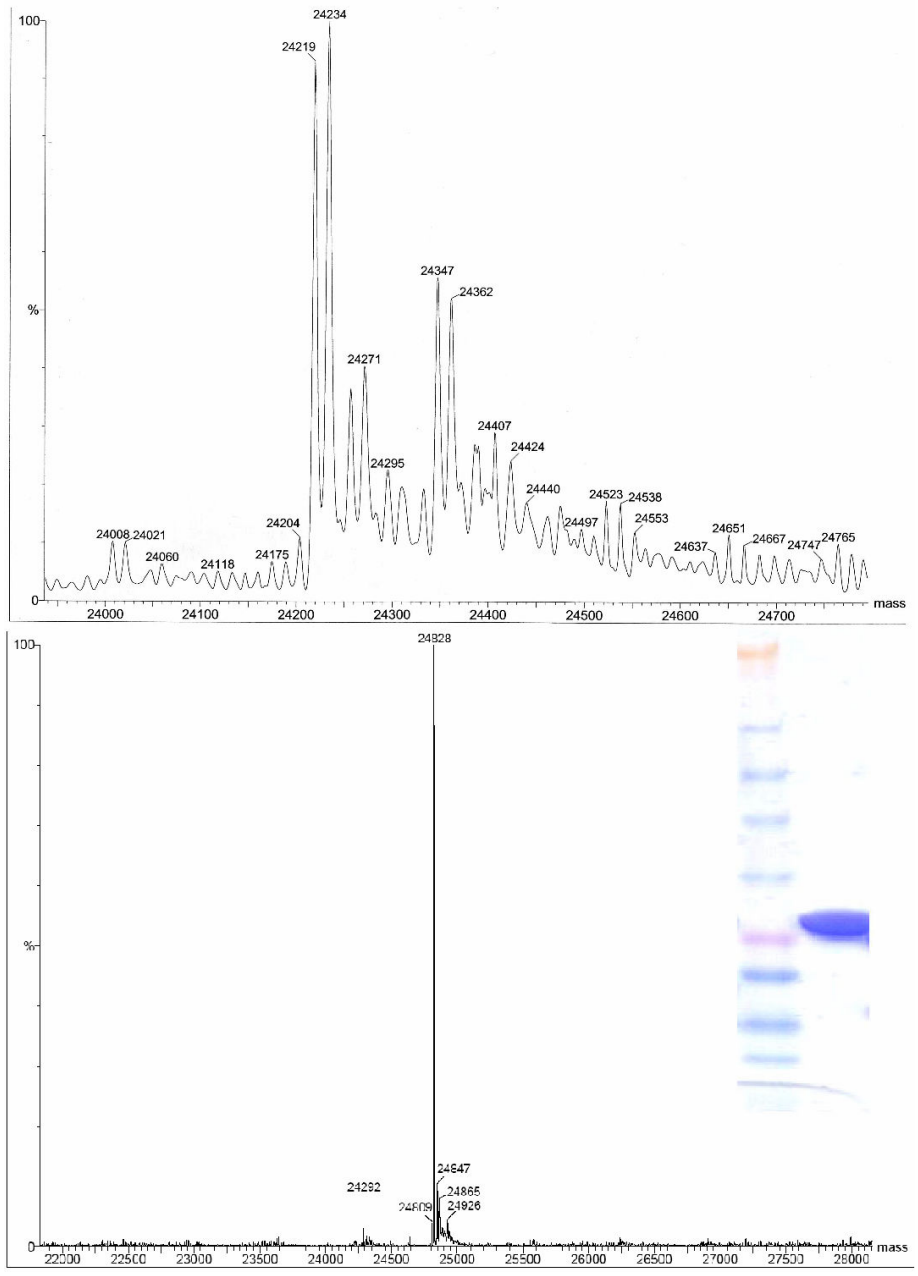

Figure 3.3: No impurities are detectable by SDS-PAGE after the MBP-fusion purification (right). ESI-Mass Spectrometry shows a substantial improvement in sample homogeneity between the wild type purification (top) and the MBP-Fusion purification (bottom).

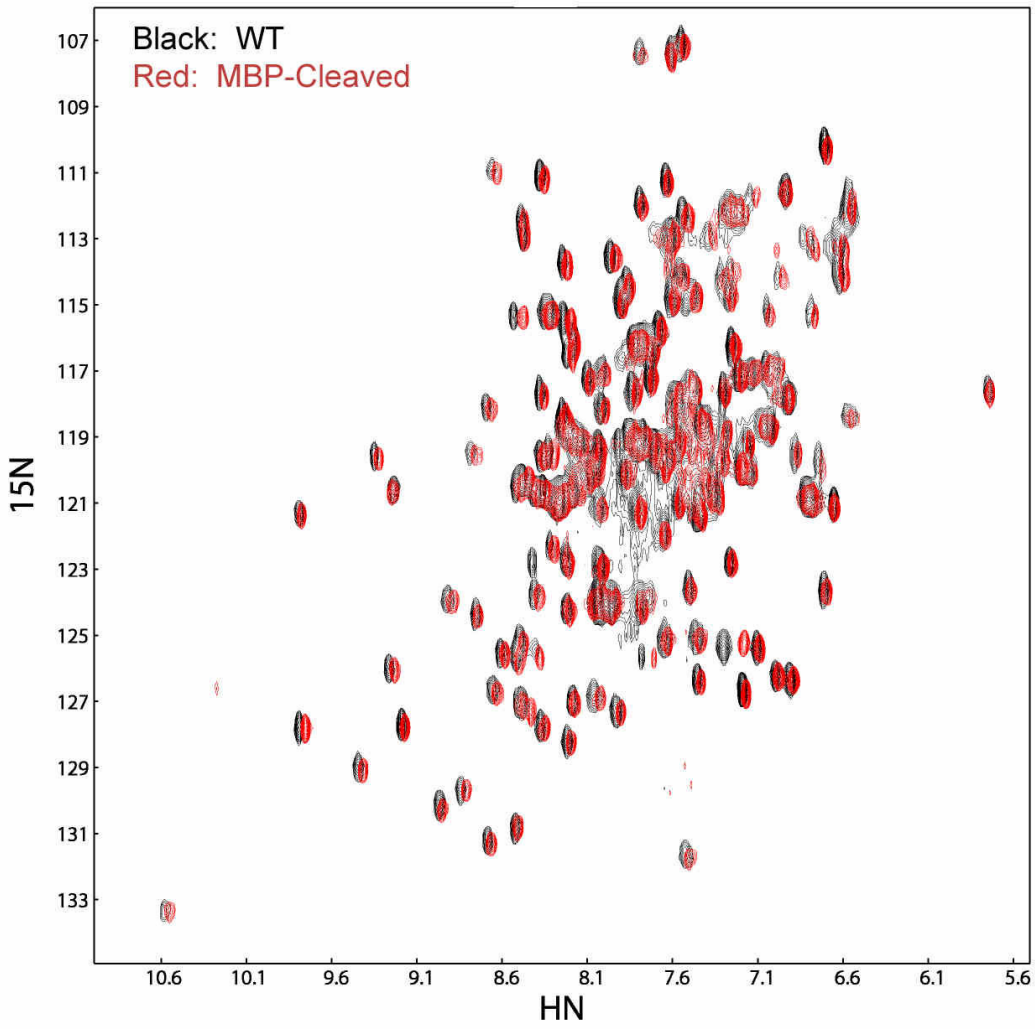

Figure 3.4: When overlaid, it is clear that the ^{15}N-HSQC spectrum of Rpp30 obtained from immobilized metal affinity chromatography (red) has much more resolution in the central region (~7-8 ppm in the HN dimension) of the spectrum than does the protein purified without the tag (black) with few other changes overall.

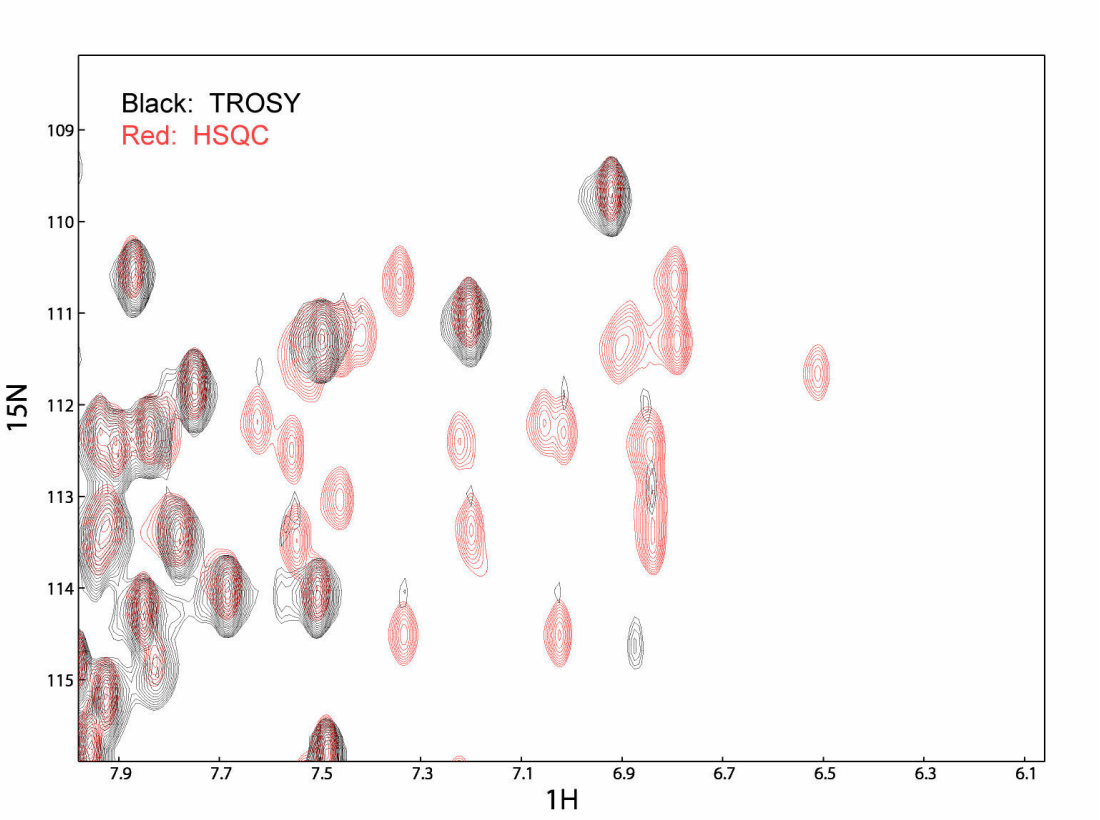

Figure 3.5: An overlay of Rpp30 ^{15}N -HSQC (red) and TROSY (black) spectra identifies side-chain peaks, which give poor signal in the TROSY experiment. The TROSY spectrum referencing has been corrected for the ½ J-coupling shift in peak position. Both spectra were collected in 20 mM KCl, 20 mM sodium acetate pH 5.0.

CHAPTER 4

INTERACTIONS BETWEEN *Pfu* Rpp30 AND *Pfu* Pop5

4.1 Introduction:

The overall aims of the Rpp30 project are to characterize its role in the RNase P holoenzyme by investigating subunit interactions. With the quality sample from Chapter 3 in hand, investigation of Rpp30 interactions was initiated using isothermal titration calorimetry (ITC) and NMR chemical shift perturbations. Rpp30 is closely associated with the RNase P protein Pop5: an interaction between the two proteins has been suggested in all yeast two-hybrid screens reported for their homologues[26,27,28], and both together can fulfill the protein requirement for a partially reconstituted RNase P (Gopalan, unpublished data). Pop5 was the obvious first target for characterizing subunit interactions of Rpp30.

ITC is a direct technique for observing the thermodynamics of molecular interactions. In an ITC experiment, a ligand or binding partner is titrated into a solution containing the molecule of interest. If the reaction is exothermic, heat is released into the solution, chiefly due to the interacting molecules achieving a lower energy state. Smaller heat-releasing effects (such as dilution of the titrant and titrand) can largely be controlled for by observing the heat generated when the titrant is titrated into a saturated solution of titrand.

The cell in which the reaction takes place is stirred constantly and, more importantly, kept at a constant temperature. When a binding event causes a change in temperature, the instrument must compensate by adjusting the amount of heat energy it is

adding to the solution. The change in power that the instrument must use to maintain the constant cell temperature is measured and used to calculate the binding constant (K_A), stoichiometry (n), and enthalpy change ΔH of a binding event by taking advantage of the isothermal environment of the reaction. With these values, the reaction's ΔG and ΔS values can be readily calculated using the basic thermodynamic formulae:

$\Delta G = - RT\ln(K)$ and $\Delta G = \Delta H - T\Delta S$.

The ultimate relation between the change of heat for the *i*-th injection (Δq_i , obtained from integrating the directly measured power with respect to time), the total concentration of titrand *M*, the concentration of titrant *L*, the volume of the cell, and thermodynamic measures comes from the equation:

$$\Delta q_i = n[M]_{\text{tot}} V_{\text{cell}} \Delta H_{\text{app}} \times \mathbb{R}$$

where $\mathbb{R}$ is the root of the equation:

$$Y_i^2 - Y_i \times \left(1 + \frac{1}{nK_A[M]_{\text{tot}}} + \frac{[L_i]_{\text{tot}}}{n[M]_{\text{tot}}}\right) + n[L_i]_{\text{tot}}[M]_{\text{tot}} = 0$$

in which Y_i is the degree of saturation: $Y_i = \Delta[L_i]\text{bound}/[M]\text{tot}$. The full derivation of this equation can be found in[65]. Data fitting using nonlinear regression allows for the quick determination of n, K_A, and ΔH from one complete titration.

NMR chemical shift perturbations are a rather more complex measurement. The chemical shift of a particular peak on an NMR spectrum has a very sensitive dependence upon the chemical environment experienced by the atom that gives rise to that peak. A binding event can cause perturbations in the chemical environment of the protein's atoms in a number of ways. Direct contact or close proximity to the ligand will cause large

changes in chemical shifts, while conformational changes in the protein can alter solvent exposure and intramolecular contacts of atoms away from the binding site.

Because it is a chemical event, the binding will have an exchange rate between free and bound states. If binding is tight and the exchange rate between the two states is slow, a peak for each present state will appear. At saturating stoichiometry, only the peaks from the bound state will be visible. If the binding is weak and exchange rate is fast, a peak will appear in between the peak positions for the free and bound states respectively via signal averaging. In the intermediate exchange range, the peaks will be averaged, but broadened to the point of indetectability. In any case but intermediate exchange, an overlay of a protein's spectrum in the presence and absence of a ligand or binding partner will reveal which peaks are experiencing a new environment due to the binding event. Of course, these data can only be translated into information about the molecules of interest if the spectral peaks are assigned to atoms in the protein.

HSQC peaks generated from protein backbone amides can be assigned by an extended analysis of several 3D spectra. The HNCO uses a magnetization transfer pulse sequence that gives signals only for protons that are covalently attached to a nitrogen that is bonded to a carbonyl carbon. For proteins, this experiment provides a three-dimensional spectrum in which peaks appear on the C-axis corresponding to the chemical shift of the backbone carbonyl carbons and the other two dimensions are defined by the N and H chemical shifts of the amide N and H from the preceding residue in the protein's primary sequence. HNCA peaks appear at points on the C-axis corresponding to the chemical shift of α–carbons, and the N and H coordinates are from the same atoms that define the HNCO's H and N coordinates. For the HNCA, however, the strongest signal

correlates an NH group to the α-carbon shift of the same amino acid (the intra-residue resonance). There is a peak that correlates NH coordinates to an α-carbon shift from the preceding residue as well (the inter-residue resonance), although this is a weaker signal because of the weaker coupling between these atoms.

The assignment process involves stringing together all of the peaks of the HNCA into a sequence; each intra-residue peak will have the same C-axis chemical shift as the inter-residue peak of the next residue in the primary sequence (**Fig 4.1**). The task is aided by the fact that each amino acid has a characteristic range of chemical shifts for its atoms. The carbon atoms of the side chain have the most varied chemical environments and therefore the most distinctive chemical shifts. By placing amino acid identities on a particular series of peaks, one can assign a string of consecutive peaks to a string of amino acids in the primary sequence of the protein.

Due chiefly to overlap issues (many peaks sharing a nearly identical carbon shift), the HNCA and HNCO alone are often not sufficient for backbone assignments in practice. But with the aid of the HNCACB (which correlates NH groups to both inter- and intra-residue α– and β–carbon shifts) and the CBCA(CO)NH (which correlates the inter-residue signals only) (**Fig 4.2**), most peaks can be assigned for a reasonably small protein with a reasonably complete set of signals.

Here we report that nearly all of the backbone resonances of Rpp30 have been assigned (**Fig 4.3**). An interaction between Rpp30 and Pop5 was observed by ITC and confirmed by large chemical shift perturbations in the NMR spectra of each protein. Mapping the chemical shift perturbations onto structural models of both proteins has suggested a binding interface for the two RNase P subunits.

4.2 Materials and Methods:

4.2.1 Preparation of Complexes

Rpp30 was prepared as described in chapter 3 and dialyzed into various buffers depending on the experiment. *Pfu* Pop5 was prepared by Ross Wilson as described earlier[29]. The C42S Pop5 mutant was used for ITC experiments. Unless otherwise noted, solutions containing both proteins were prepared by injecting a Pop5 solution into a solution containing Rpp30. NMR samples contained a roughly 1:1 ratio of proteins with only one of the proteins ^{15}N labeled per sample. Concentrations of 125 μM each were used for the ^{15}N-Rpp30 spectrum and 260 μM Pop5 with 280 μM Rpp30 were used in the ^{15}N-Pop5 spectra. Mixtures of Rpp29 and Rpp30 were 200 μM and were formed by titrating one equivalent of unlabeled Rpp29 into ^{15}N Rpp30.

4.2.2 ITC

Unlabeled GEF-Rpp30 and C42S Pop5 were diluted to 20 μM and 200 μM respectively. Proteins were dialyzed together against the same buffer containing only 5 mM sodium acetate and 0.01% NaN_3 pH 5.0, and each sample was degassed for 3 min at 25 ^oC. The Pop5 sample was loaded into the syringe and injected into the Rpp30 solution in the cell; the first injection was 3 μL and subsequent injections were 5 μL with a delay of 220 s between injections. The heat of injection was estimated using the non-reacting (last 4) injections. Data were fit with Origin software (MicroCal).

4.2.3 NMR Data Collection

Samples used for NMR experiments were either unlabeled (grown in LB-broth) or uniformly labeled with either ^{15}N or both ^{15}N and ^{13}C (grown in M9 minimal media supplemented with 1% Eagle Basal Vitamin Mix (Life Technologies, Gaithersburg,

MD)[57], containing 1 g/liter ^{15}N ammonium chloride as the sole nitrogen source and 3 g/liter ^{13}C glucose as the sole carbon source for ^{13}C labeled samples). After dialysis into the appropriate buffer, samples were concentrated to 500 μL and D_2O was added to 5%.

3D experiments were collected from a 1 mM sample in a buffer containing only 10mM sodium acetate, 0.01% NaN_3, pH 5.0 at 55 ^oC on a Bruker DRX-600 spectrometer equipped with a cryogenic probe. 2D ^{15}N HSQC experiments of the Rpp30-Pop5 complex were collected in 15mM sodium acetate, 0. 01% NaN_3, pH 5.0 with one uniformly ^{15}N labeled protein and one unlabeled protein. Data was collected on a Bruker DMX-800 spectrometer at 55 ^oC. Data were processed with NMRPipe[58] and NMRView[59].

4.3 Results and Discussion:

4.3.1 Rpp30-Pop5 interaction by ITC

ITC data collected on Pop5 titration into Rpp30 are less than ideal. However, a distinct binding curve of roughly 1-to-1 stoichiometry was observed, indicating a moderately tight binding event between the two proteins even at 25 ^oC (**Fig 4.4**). Although the calculated K_D of ~900 nM may not be relevant to the true nature of the holoenzyme, the ITC titration gives strong evidence for 1-to-1 binding between Rpp30 and Pop5 in solution.

4.3.2 Rpp30-Pop5 interaction by NMR

A distinct interaction between Rpp30 and Pop5 in solution was also detected by NMR; major chemical shift perturbations in each protein's ^{15}N HSQC spectrum were observed upon addition of the other protein. To observe shifts in the Rpp30 spectrum, an HSQC spectrum was recorded on ^{15}N labeled protein in the absence and presence of

unlabeled Pop5. Comparing the spectra revealed changes in a large number of peak positions, revealing extensive interactions between the proteins (**Fig 4.5**). The spectral resolution for the sample containing both proteins was comparable to that of isolated Rpp30, suggesting that the interaction was specific and tight; non-specific interactions would instead produce a heterogeneous sample, degrading the signal substantially.

Similar results were observed with the same experiment using ^{15}N Pop5. Interestingly, when Rpp30 was titrated into a Pop5 solution, the spectrum was all but annihilated (**Fig 4.6**); the reverse titration (Pop5 into Rpp30) gave full spectra (**Fig 4.7**). We suspect that this phenomenon was due to nonspecific aggregation that took place only when large quantities of Pop5 are present with limited amounts of Rpp30, perhaps as a byproduct of dimerization of the complex.

4.3.3 Assignment of the Rpp30 Backbone

To interpret the chemical shift changes observed in the Rpp30 spectrum upon the formation of the Rpp30-Pop5 complex, the Rpp30 backbone residues were assigned. HNCO, HNCA, HNCACB, and CBCA(CO)NH spectra were sufficient to assign 191 residues (**Fig 4.3**) thanks to a robust sample signal aided by a cryogenic probe. During this time an HCCCONH experiment was also recorded, but the spectral quality was rather poor and its data were used sparingly. Discounting the 14 proline residues and the N-terminal glycine, the final assigned peak list represents ~95% of the assignable residues. These backbone assignments will be critical for any further NMR study of *Pfu* Rpp30 and paves the way to numerous possible experiments for elucidating its role in RNase P, including structural characterization.

A prediction of the protein's secondary structure was calculated based on the chemical shifts of each residue using the CSI database comparison[66] (**Fig 4.8**). The predicted secondary structure of Rpp30 in solution is largely consistent with the alternating helix-sheet pattern observed in the crystal of the *Pho* protein. What little deviation can be observed occurs mostly at the N- and C- termini of the protein and around residue 100 where NMR data predicts a loop in place of the crystallographically observed β-5 strand; none of these three areas is highly conserved.

4.3.4 Mapping Chemical Shift Perturbations

Ideally, Rpp30's peaks would be completely assigned in both the free and Pop5-bound states; in that situation, the degree of chemical shift change could be measured for each residue and mapped onto a sequence and/or structure with the largest chemical shift changes revealing sites of interaction between the two proteins. Substantial information about the ineraction can also be achieved using only the available assignments for free Rpp30. In this case, the spectrum of the complex could be partially reassigned using a quick and inexpensive HSQC experiment. Non-overlapping peaks that did not shift at all were easily reassigned, and in the peripheral regions of the spectrum a number of small and medium shift perturbations could be unambiguously reassigned. Although the shifts of many residues in the protein still could not be determined using this technique, the preliminary results of the 2D experiment are still informative.

The magnitude of the measured chemical shift perturbation was mapped onto the primary sequence of Rpp30. Using a homology model of the protein based on the highly homologous *Pho* Rpp30 crystal structure (model generated using SWISS-MODEL: http://swissmodel.expasy.org//SWISS-MODEL.html), the magnitudes could be mapped

onto a three-dimensional model as shown in **Fig 4.9**. The largest shifts were generally localized to one face of the protein, with moderate shifts observed in the central pore. This suggests that Rpp30 binds Pop5 on the region surrounding helix-10 with some degree of conformational change.

Data from an HNCO of the complex with unlabeled Rpp30 and uniformly ^{15}N and ^{13}C labeled Pop5 allowed Ross Wilson to reassign ~60% of the Pop5 peaks, and mapping the chemical shift changes revealed a likely reciprocal interface on the Pop5 crystal structure (**Fig 4.10**). Our results are consistent with the hypothesis that Pop5 (which adopts the RNA Recognition Motif fold) binds to RNA in a similar fashion as the bacterial RNase P protein: Rpp30 binds to a face that would not be involved with the suspected RNA-binding. The binding interface identified for each protein in solution is consistent with a recently published crystal structure of the Rpp30-Pop5 complex from *Pho* (**Fig 4.11**). The crystal structure showed only a 0.6 Å RMS deviation between Rpp30 free and bound, indicating almost no structural change upon binding. However, even small changes in conformation could translate into significant alterations in chemical environment for the protein's constituent atoms.

In contrast to the results observed upon with addition of Pop5, titrating one equivalent of unlabeled Rpp29 into ^{15}N-Rpp30 produced almost no spectral perturbations (**Fig 4.12**). What few (< 10) peaks that showed minor movement were unassigned. Such minor changes in the spectrum are inconsistent with an intimate interaction between these proteins under the given conditions.

4.3.5 Conclusions and Future Directions

Identification of a Pop5-Rpp30 interface is an important step toward understanding how each protein is involved in the holoenzyme, and we now have reason to suspect that there is no specific Rpp29-Rpp30 interaction, which will simplify future experiments involving both proteins. The door is open to investigate other Rpp30 interactions on the way to understanding its functional role.

Particularly, it will be interesting to see how the protein is interacting with the catalytic RNA subunit and the substrate tRNA to elucidate the role of protein-RNA interactions in the holoenzyme. Electrophoretic mobility shift assays involving Rpp30 or the Rpp30-Pop5 complex combined with an RNA subunit or truncated version thereof will be the first step in characterizing the relationship between protein and RNA subunits. On top of chemical shift perturbations, there are a variety of NMR techniques available that may be used advance our understanding of the interactions in more detail, including experiments that measure H/D exchange, residual dipolar couplings, molecular cross-saturation, or the full solution structure of various Rpp30 complexes. A full structure of the holoenzyme would likewise be informative, and crystallographic work toward that end is already underway.

In addition, a more detailed description of the Rpp30-Pop5 interaction will inform experiments involving the holoenzyme. For instance, more ITC data would provide the K_D of complex formation, which is an important parameter for future reconstitution and crystallization efforts. Also, the binding interface could be defined with greater precision by recording 3D spectra of labeled Rpp30 in complex with Pop5, which would allow

further site-directed mutagenesis experiments to be conducted, probing the significance of the interaction.

In all, these are small but necessary steps in coming to a better understanding of RNase P, which is itself a small but important component of every cell's existence.

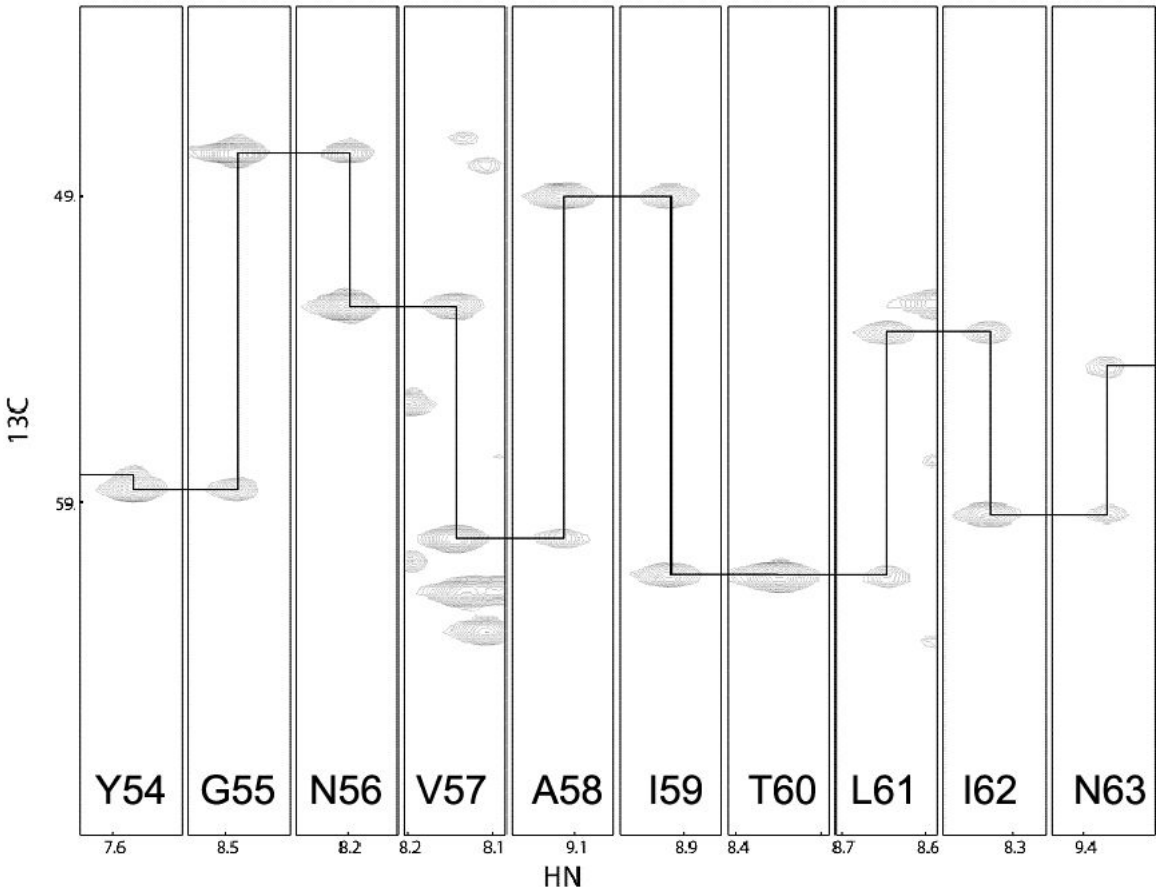

Figure 4.1: Strip plots from the Rpp30 HNCA spectrum demonstrate how to relate NMR peaks of a particular residue to its relative location in the primary sequence. Each strip displays a small ppm range for the ^{15}N and HN dimensions and contains two different peaks in the ^{13}C (vertical) dimension, which represent the resonances from the inter-residual and intra-residual α-carbons. Identical carbon shifts can be found for adjacent residues, linking neighboring amino acids. The strips are labeled with the identity of the corresponding residue.

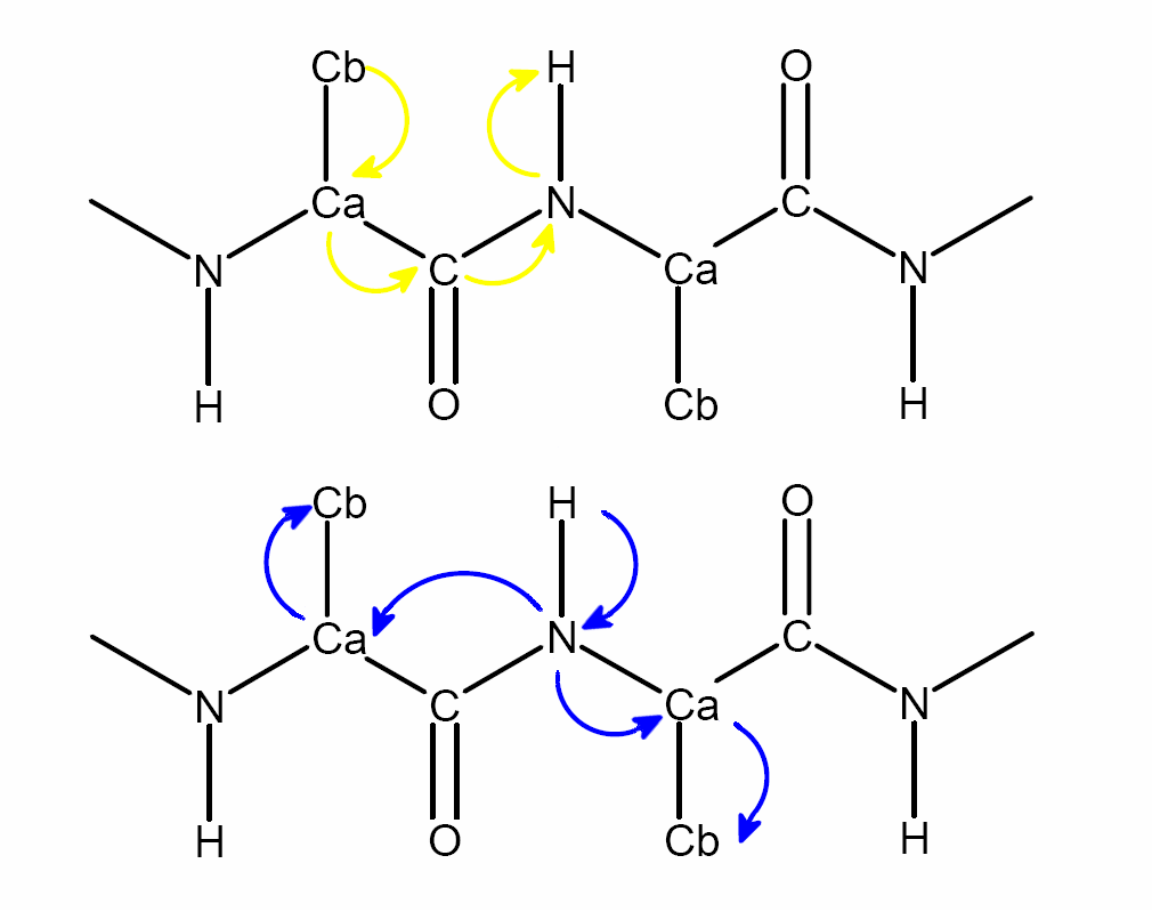

Figure 4.2: Using spin-echo pulse sequences to selectively transfer nuclear excitation between atoms is crucial for NMR characterization of proteins. Above are the pathways of magnetization transfer along the backbone of the protein for the CBCA(CO)NH (yellow) and HNCACB (blue) experiments. The CBCA(CO)NH relates an amide proton to the sidechain carbons of the preceding residue, whereas the HNCACB relates the amide proton to both inter-residual and intra-residual sidechain carbons.

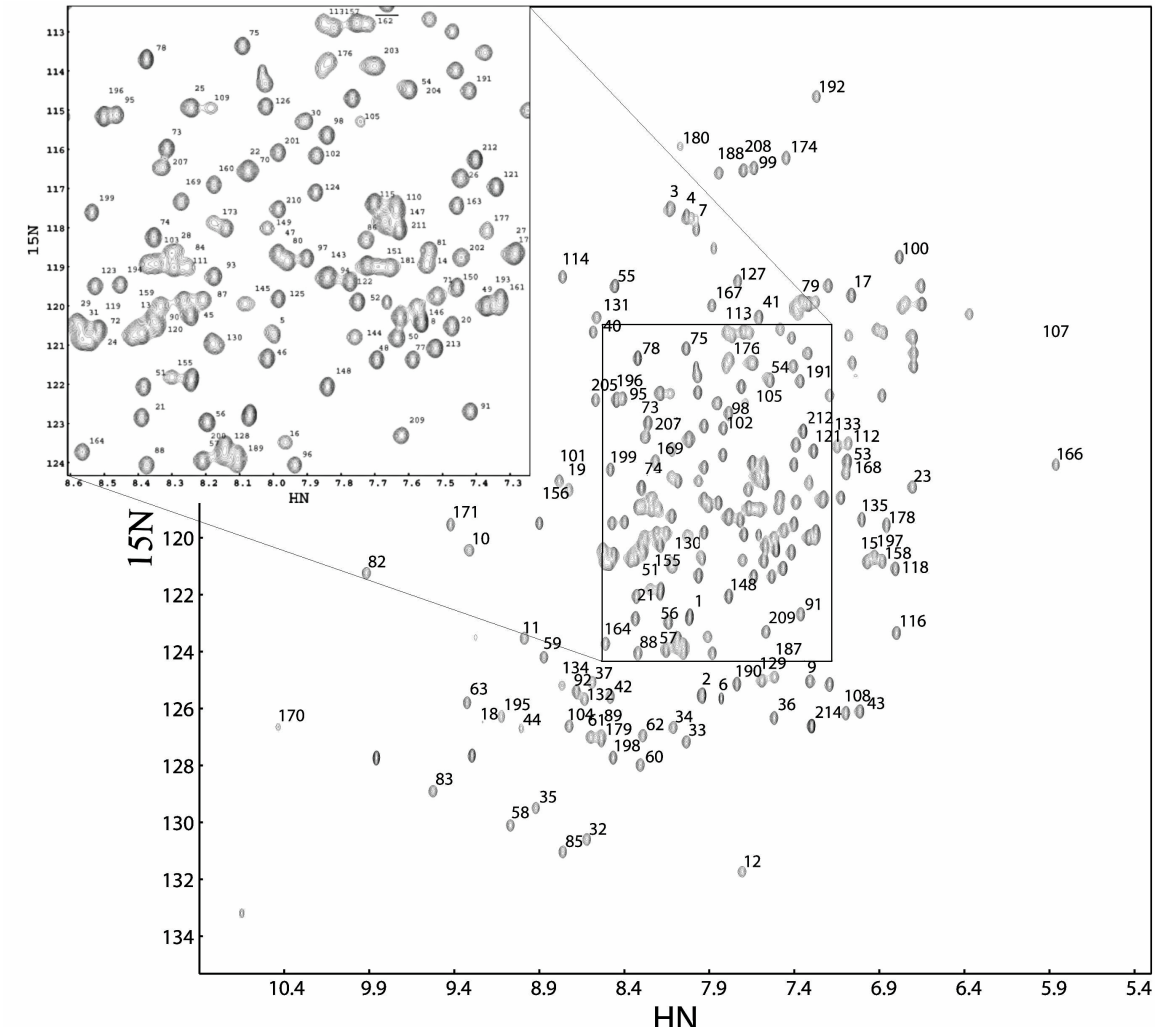

Figure 4.3: Peaks of the Rpp30 HSQC are labeled according to backbone assignments. The number corresponding to the residue's position in the primary sequence appears to the upper-right of each assigned peak. The central region of the spectrum has been expanded in the upper left for clarity.

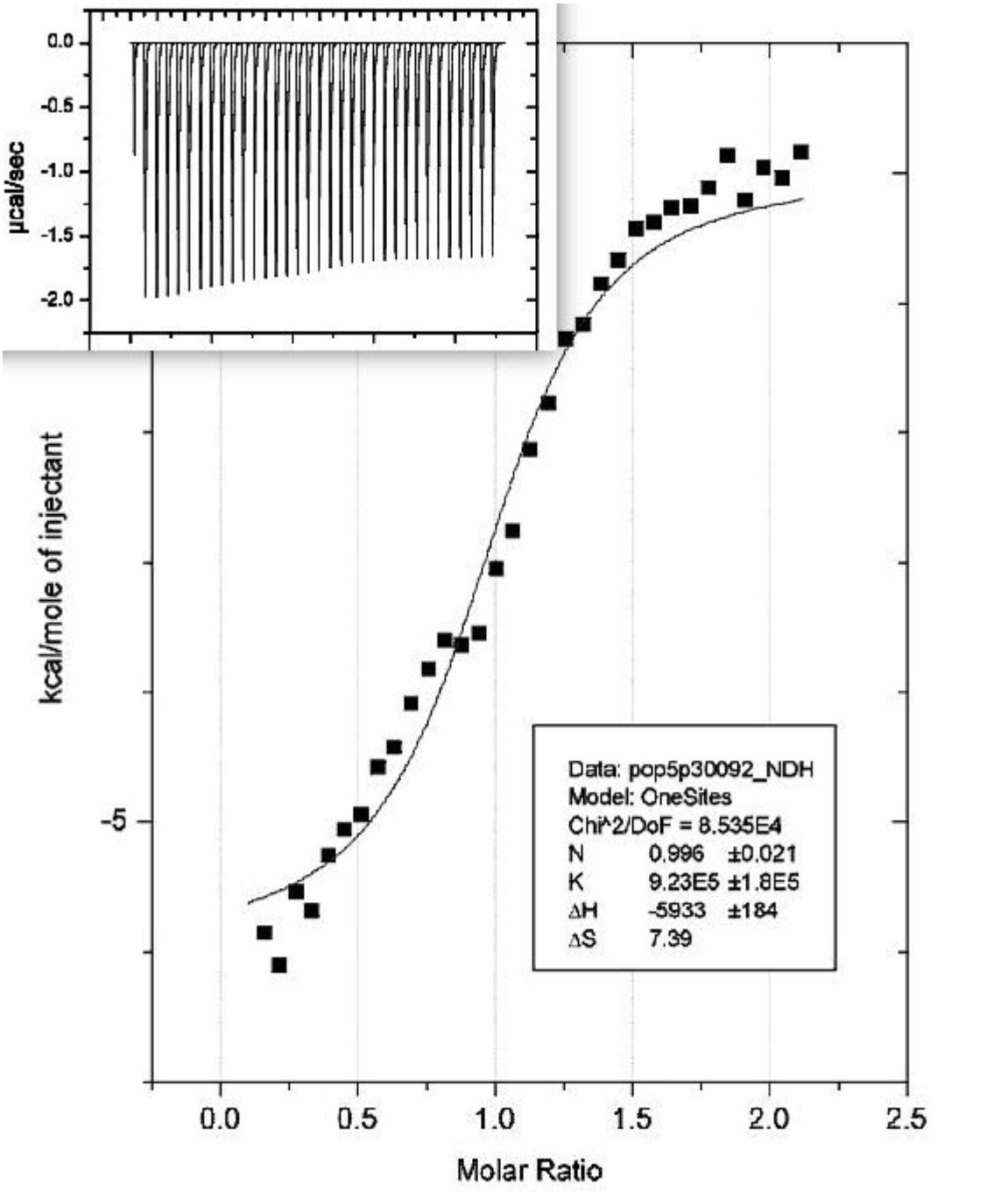

Figure 4.4: ITC curve of Pop5 (21 µM protein) titrated into Rpp30 (210 µM protein) shows a sigmoidal curve characteristic of a one-to-one binding event. The data was collected at 25 ^{o}C in 10 mM sodium acetate, pH 5.0.

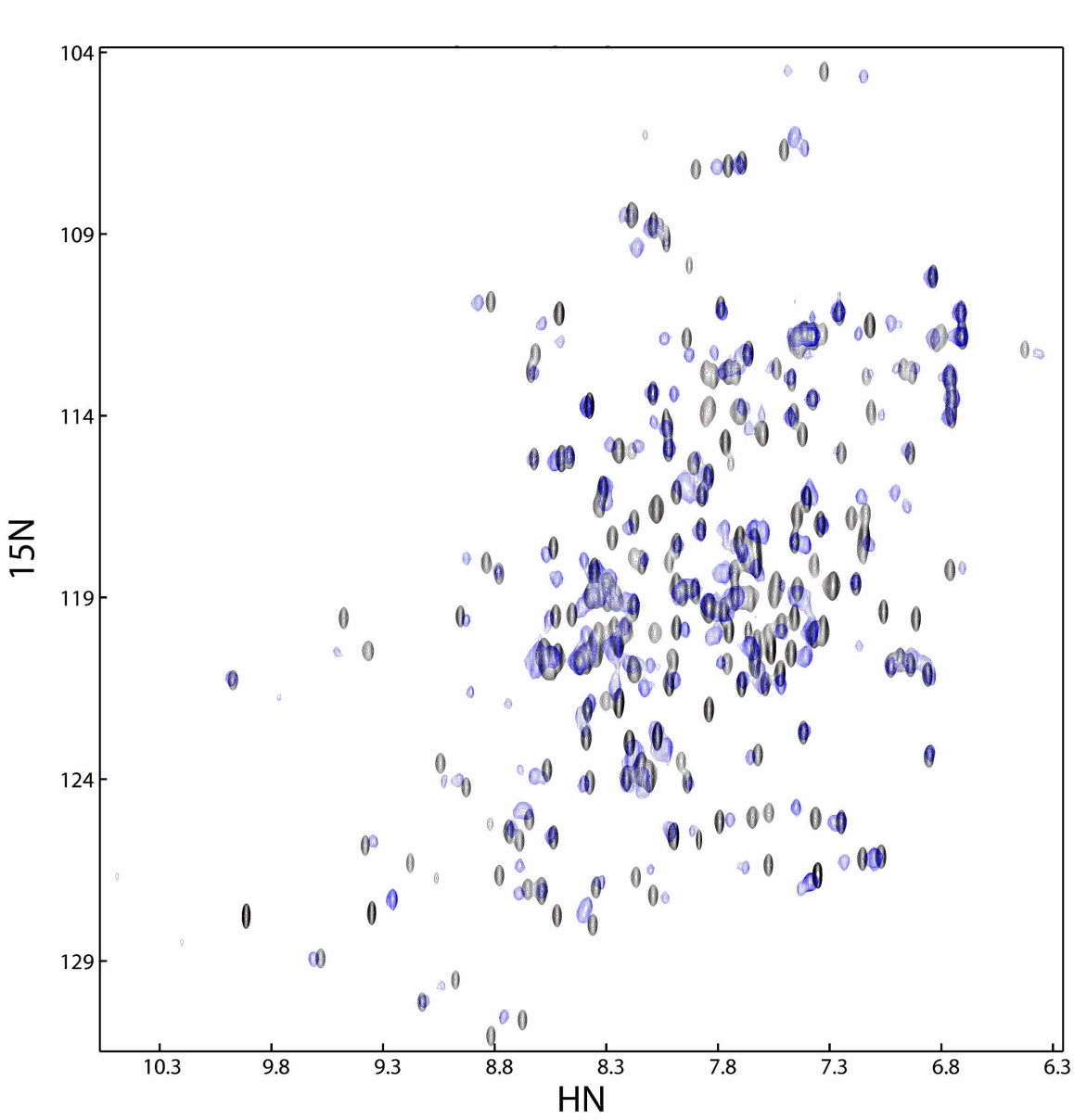

Figure 4.5: An overlay of HSQC spectra for free Rpp30 (black) and Rpp30 bound to Pop5 present in 1:1 stoichiometry (blue) shows that a large number of Rpp30's amide protons experience a different chemical environment in the absence and presence of Pop5. Such extensive effects are likely to be caused by specific protein-protein interactions in the given condition of 10 mM sodium acetate, pH 5.0.

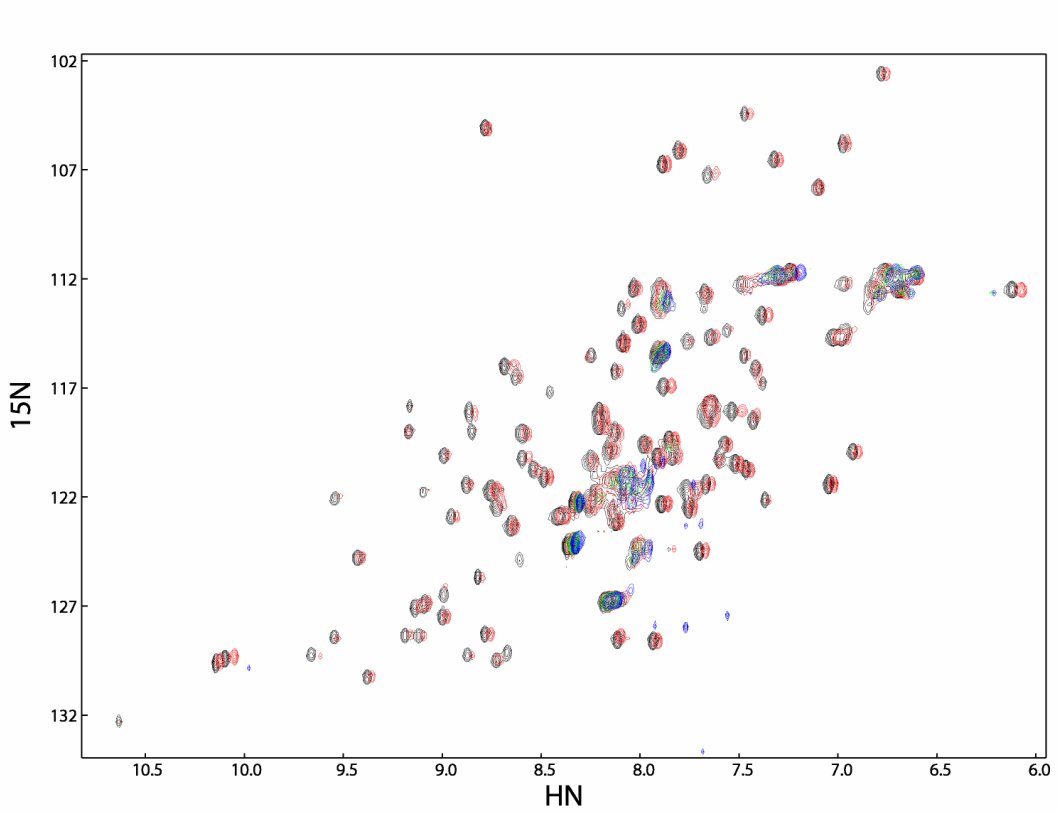

Figure 4.6: Titration of Rpp30 into ^{15}N-labeled Pop5 abolishes the spectral resolution as is observable by the disappearance of the majority of peaks as the titration proceeds. The spectrum of free Pop5 (black) is overlaid with the following Pop5-to-Rpp30 stoichiometries: 4:1 (red), 2:1 (green), 1:1 (blue). This result is likely caused by aggregation of the complex or nonspecific binding that occurs when few Rpp30 molecules are available in an excess of Pop5.

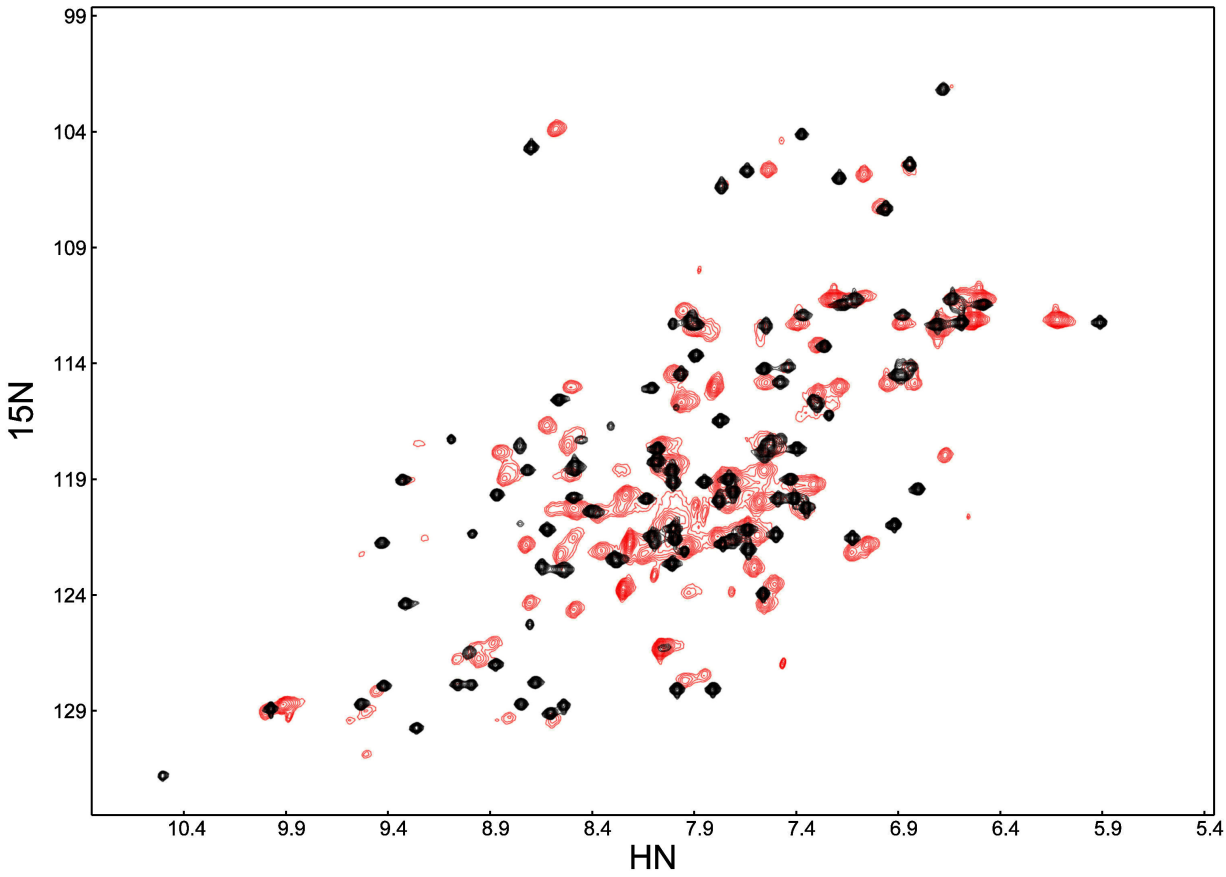

Figure 4.7: An overlay of HSQC spectra for free Pop5 (black) and Pop5 bound to Rpp30 in 1:1 stoichiometry (red) when Pop5 is titrated into an Rpp30 solution again shows major spectral changes indicative of a specific interaction between the two proteins. It is also notable that the spectral quality of Pop5 is reduced when bound to Rpp30, which could be caused by increased relaxation (due to increased hydrodynamic radius) or heterogeneity in the complex interface.

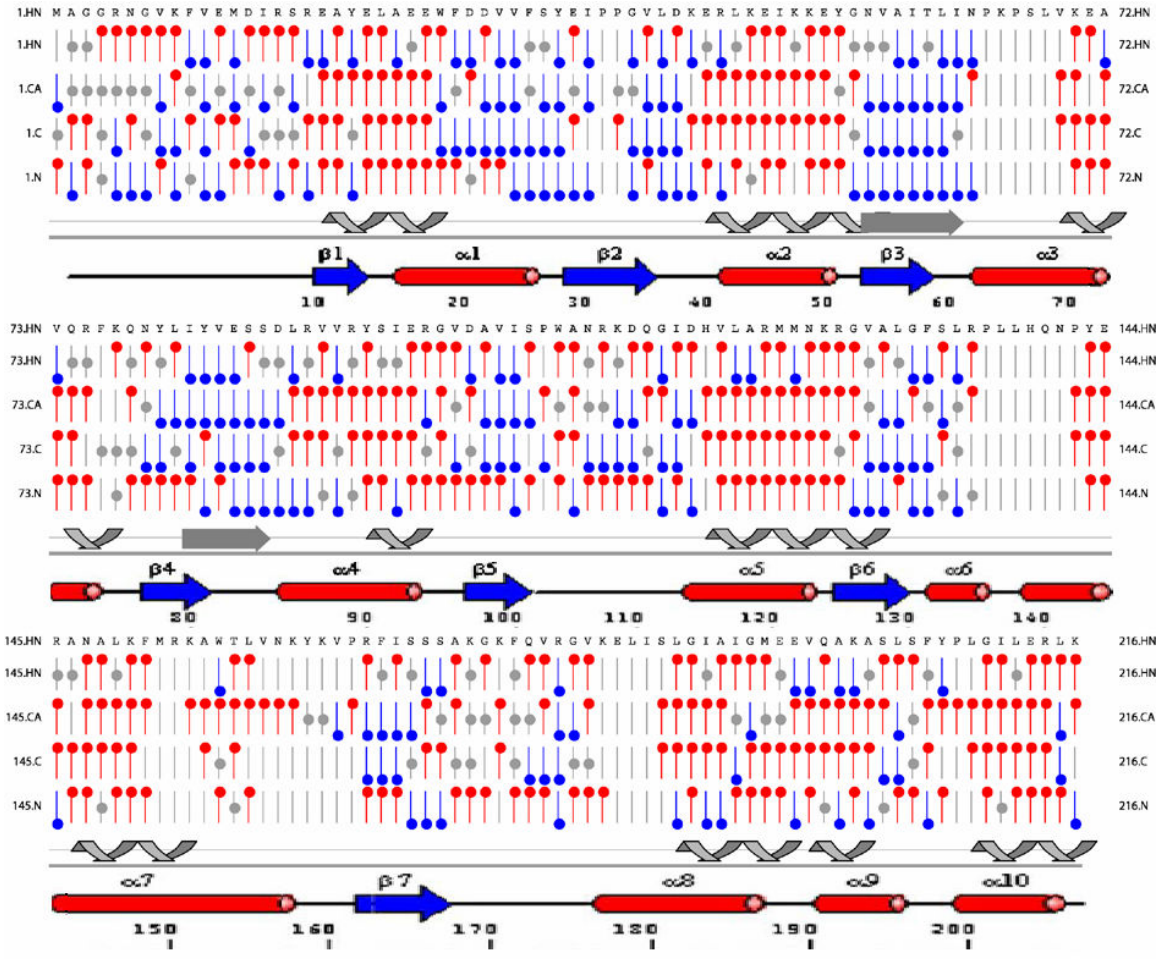

Figure 4.8: CSI secondary structure prediction for *Pfu* Rpp30 is compared to that of the *Pho* Rpp30 crystal structure. Rows of red and blue dots represent helical and strand-like chemical shifts respectively for (from top to bottom) amide protons, alpha carbons, carbons, and nitrogens of each backbone residue. The cartoon in gray under each segment is a predicted structural element based on a computation using each atom[66]. The cartoon in color represents the secondary structure of *Pho* Rpp30 described in [35], whose 211 residue sequence can be aligned with *Pfu* Rpp30's almost exactly.

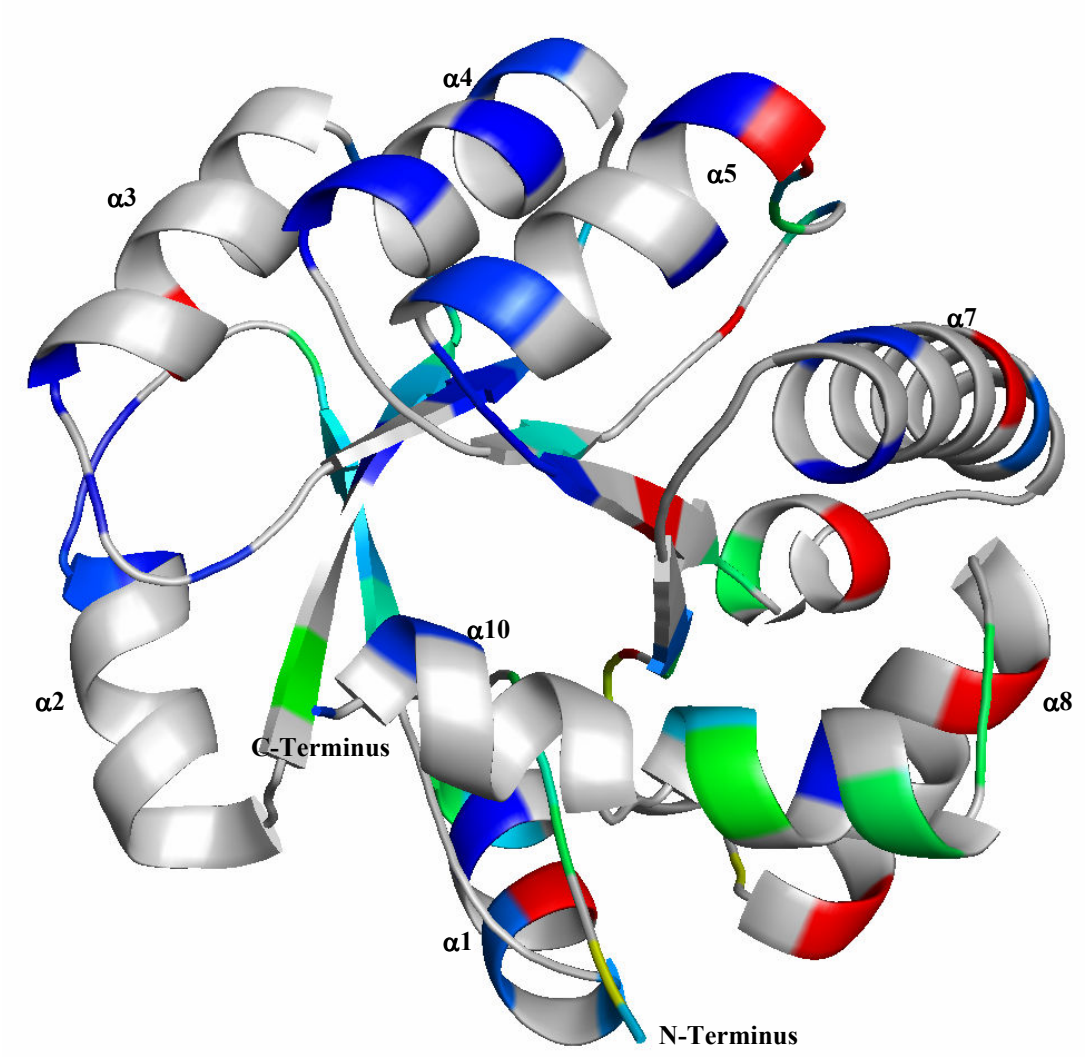

Figure 4.9: Magnitudes of peak shifts were mapped onto this homology model of *Pfu* Rpp30 based on the *Pho* crystal structure. Although much of the protein (gray) was not reassigned, the color gradient from blue (no change) to red (big change) shows that most of the identifiable peak movements were localized to one side of structure (the right in this figure).

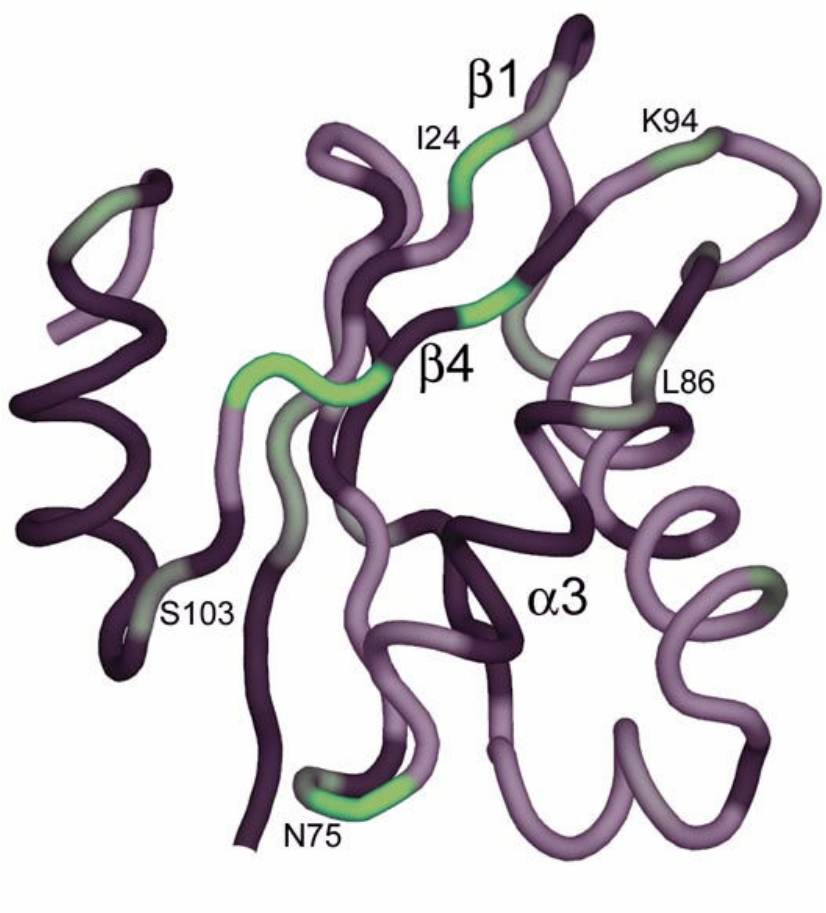

Figure 4.10: Pop5 residues labeled with a color ramp from gray to green to indicate minor to major chemical shift perturbations respectively show that major NMR peak movements are generally localized to one face (left side of figure) of the protein's structural model. Black residues correspond to amino acids whose chemical shift change could not be determined unambiguously from the HNCO.

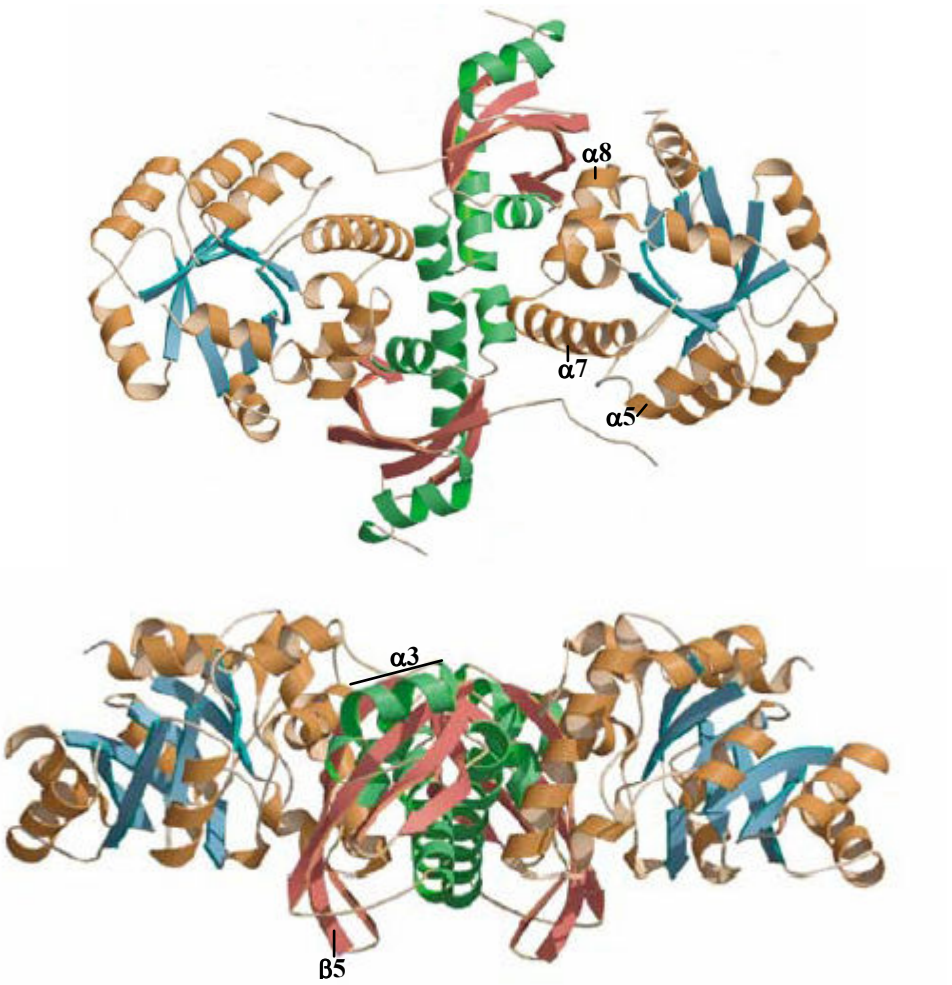

Figure 4.11: The Rpp30-Pop5 interface identified by NMR chemical shift changes corresponds well to a recently published crystal structure of the complex from *Pho*[34], shown here in two orientations. The crystal comprises four subunits with two Rpp30 monomers (gold and blue) flanking two interacting Pop5 monomers (green and red). The secondary structural elements most involved in the interface are labeled for Rpp30 in the top figure and for Pop5 in the bottom figure.

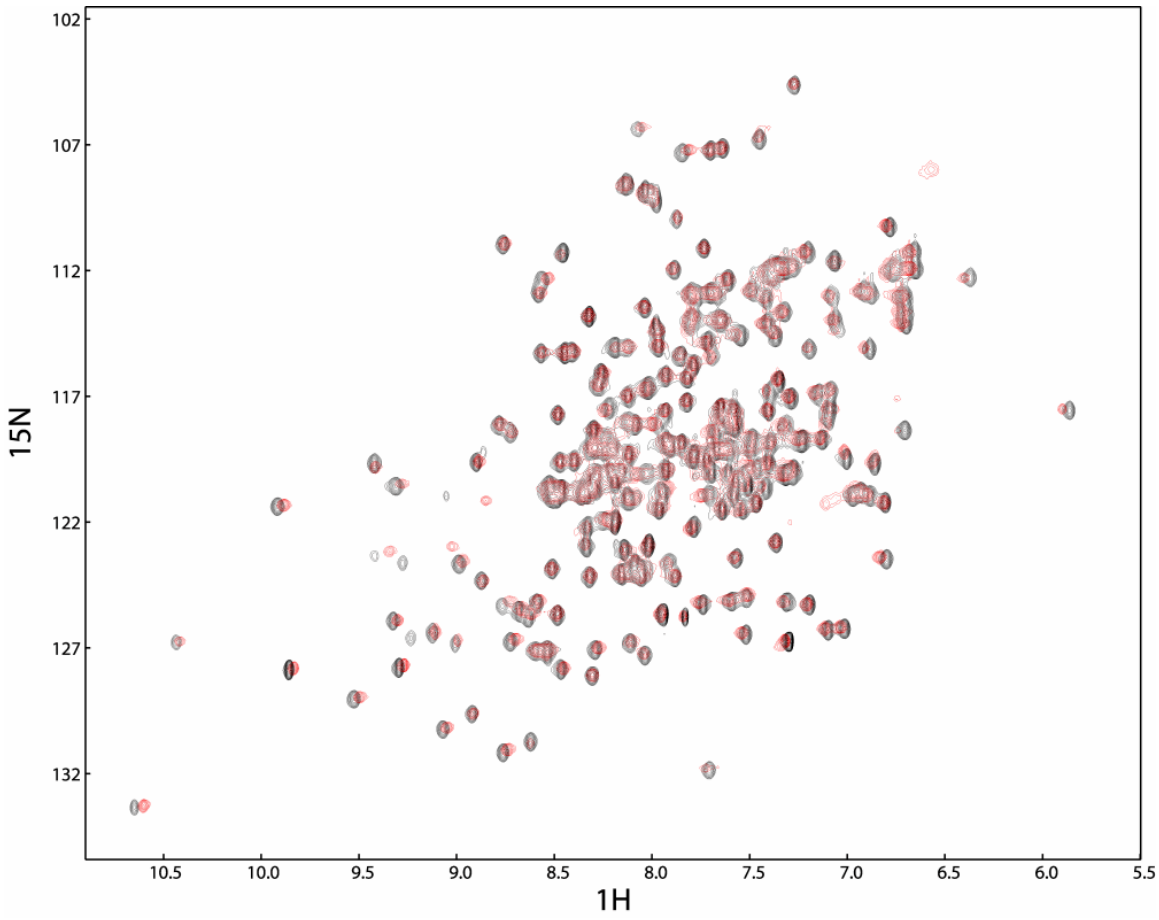

Figure 4.12: Overlaying HSQC spectra for Rpp30 free (black) and in the presence of one equivalent of Rpp29 (red) reveals almost no spectral changes. There is therefore not a major interaction between the two proteins in the condition tested.

References:

[1]Altman, S. Enzymatic Cleavage of RNA by RNA, *Nobel Lectures, Chemistry 1981-1990*, Ed. Malmström, B.G. (World Scientific Publishing Co., Singapore, 1992).

[2] Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. (1983) *Cell* 35, 849-857.

[3] Altman, S. & Kirsebom, L. (1999) in *The RNA World*, eds. Gesteland, R. F., Cech, T. R. & Atkins, J. F., pp. 351-380.

[4] Robertson, H.D., Altman, S., Smith, J.D. (1972) *The Journal of Biological Chemistry* 247, 16, 5243-51.

[5] Harris, M. E., Frank, D. N. & Pace, N. R. (1998) Structure and Function of the Bacterial Ribonuclease P Ribozyme. In *RNA Structure and Functions* eds Simons, R. W. & Grunberg-Manago, M. (Cold Spring Harbor Laboratory Press, New York)pp. 309-38.

[6] Frank, D.N. Pace, N.R. (1998) *Annu. Rev. Biochem*. 67, 153–180.
Paul, R., Lazarev, D., and Altman, S. (2001) *Nucleic Acids Res*. 15; 29(4): 880–885.

[7] Paul, R., Lazarev, D., and Altman, S. (2001) *Nucleic Acids Res*. 15; 29(4): 880–885.

[8] Darr SC, Brown JW, Pace NR. (1992*) Trends Biochem Sci.*17(5):178-82.

[9] Gopalan, V., Baxevanis, A. D., Landsman, D. & Altman, S. (1997) *J. Mol. Biol.* 267, 818-29.

[10] Niranjanakumari, S., Stams, T., Crary, S. M., Christianson, D. W. & Fierke, C. A. (1998) *Proc. Natl. Acad. Sci. U S A* 95, 15212-7.

[11] Kurz, J. C., Niranjanakumari, S. & Fierke, C. A. (1998) *Biochemistry* 37, 2393-400.

[12] Kurz, J.C. and Fierke, C.A. (2002) *Biochemistry*, 41, 30, 9545 – 9558.

[13] Schedl, P. & Primakoff, P. (1973) *Proc. Natl. Acad. Sci. U S A* 70, 2091-5.

[14] Buck AH, Dalby AB, Poole AW, Kazantsev AV, Pace NR. (2005) *The EMBO Journal* 24**,** 3360–3368.

[15] Hartmann, E., Hartmann, R. (2003) *TRENDS in Genetics* 19;10: 561-69

[16] Brown, J.W. (1999) *Nucleic Acids Res*. 1;27(1):314.

[17] Forster, A. C. & Altman, S. (1990) *Cell* 62**,** 407–409.

[18] Xiao, S., Houser-Scott, F. & Engelke, D. R. (2001) *J. Cell. Physiol.* 187**,** 11–20.

[19] Harris JK, Haas ES, Williams D, Frank DN, Brown JW. (2001) *RNA.* 7(2):220-32.

[20] Pannucci, J. A., Haas, E. S., Hall, T. A., Harris, J. K. & Brown, J. W. (1999) *Proc. Natl. Acad. Sci. USA* 96**,** 7803–7808.

[21] Boomershine, W. P., McElroy, C. A., Tsai, H. Y., Wilson, R. C., Gopalan, V. & Foster, M. P. (2003) *Proc. Natl. Acad. Sci. USA* 100**,** 15398–15403.

[22] Kouzuma, Y., Mizoguchi, M., Takagi, H., Fukuhara, H., Tsukamoto, M., Numata, T. & Kimura, M. (2003) *Biochem. Biophys. Res. Commun.* 306**,** 666–673.

[23] Fukuhara H, Kifusa M, Watanabe M, Terada A, Honda T, Numata T, Kakuta Y, Kimura M. (2006) *Biochem Biophys Res Commun*. 343(3):956-64.

[24] Hall, T. A. & Brown, J. W. (2002) *RNA* 8**,** 296–306.

[25] Joyce, G. F. & Orgel, L. E. (1999) in *The RNA World*, eds. Gesteland, R. F., Cech, T. R. & Atkins, J. F. (Cold Spring Harbor Laboratory Press, Plainview), pp. 49-77.

[26] Kifusa, M., Fukuhara, H., Hayashi, T. & Kimura, M. (2005) *Biosci. Biotechnol. Biochem.* 69**,** 1209–1212.

[27] Hall, T. A. & Brown, J. W. (2004) *Archaea* 1**,** 247–253.

[28] Houser-Scott F., Xiao S., Millikin C. E., Zengel J. M., Lindahl L., Engelke D. R.. (2002) *Proc Natl Acad Sci U S A.* 5;99(5):2684-9.

[29] Wilson R. C., Bohlen C. J., Foster M. P., Bell C. E. (2006) *Proc Natl Acad Sci U S A* 103(4):873-8.

[30] Sidote, D. J. & Hoffman, D. W. (2003) *Biochemistry* 42**,** 13541–13550.

[31] Sidote, D. J., Heideker, J. & Hoffman, D. W. (2004) *Biochemistry* 43**,** 14128–14138.

[32] Numata, T., Ishimatsu, I., Kakuta, Y., Tanaka, I. & Kimura, M. (2004) *RNA* 10**,** 1423–1432

[33] Kakuta, Y., Ishimatsu, I., Numata, T., Kimura, K., Yao, M., Tanaka, I. & Kimura, M. (2005) *Biochemistry* 44**,** 12086–12093.

[34] Kawano S., Nakashima T., Kakuta Y., Tanaka I., Kimura M. (2006) *J Mol Biol* 24;357(2):583-91.

[35] Takagi, H., Watanabe, M., Kakuta, Y., Kamachi, R., Numata, T., Tanaka, I. & Kimura, M. (2004) *Biochem. Biophys. Res. Commun.* 319**,** 787–794.

[36] Hsieh J., Andrews A.J., Fierke C.A. (2004) *Biopolymers* 73, 1; 79-89

[37] Sharin E., Schein A., Mann H., Ben-Asouli Y., Jarrous N. (2005) *Nucleic Acids Res.* 9;33(16):5120-32.

[38] Trang P., Kim K., Liu F. (2004) *Cell Microbiol*. 6(6):499-508.

[39] Forster A. C., Altman S. (1990) *Cell.* 62(3):407-9.

[40] Schlegl J, Furste J. P., Bald R., Erdmann V. A., Hartmann R. K. (1992) *Nucleic Acids Res.* 25;20(22):5963-70.

[41] Forster AC, Altman S. (1990) *Science*. 249(4970):783-6.

[42] Massire, C., Jaeger, L. & Westhof, E. (1998) *J. Mol. Biol.* 279, 773–793.

[43] Tsai, H. Y., Masquida, B., Biswas, R., Westhof, E. & Gopalan, V. (2003) *J. Mol. Biol.* 325**,** 661–675.

[44] Torres-Larios, A., Swinger, K. K., Krasilnikov, A. S., Pan, T. & Mondragon, A. (2005) *Nature* 437**,** 584–587.

[45] Kazantsev, A. V., Krivenko, A. A., Harrington, D. J., Holbrook, S. R., Adams, P. D. & Pace, N. R. (2005) *Proc. Natl. Acad. Sci. USA* 102, 13392–13397.

[46] Stams, T., Niranjanakumari, S., Fierke, C. A. & Christianson, D. W. (1998) *Science* 280, 752–755.

[47] Spitzfaden, C., Nicholson, N., Jones, J. J., Guth, S., Lehr, R., Prescott, C. D., Hegg, L. A. & Eggleston, D. S. (2000) *J. Mol. Biol.* 295, 105–115.

[48] Kazantsev, A. V., Krivenko, A. A., Harrington, D. J., Carter, R. J., Holbrook, S. R., Adams, P. D. & Pace, N. R. (2003) *Proc. Natl. Acad. Sci. USA* 100, 7497–7502.

[49] Buck AH, Kazantsev AV, Dalby AB, Pace NR. (2005) *Nat Struct Mol Biol*. 12(11):958-64.

[50] Li Y, Altman S. (2002) *Nucleic Acids Res.* 1;30(17):3706-11.

[51] Dlakic, M. (2005) *RNA*. 11(2):123-7.

[52] Parella, T. (1998) "*eNMR"* in *The BRUKER NMR Guide & Encyclopedia V3.0*, http://www.bruker.de/guide/index.html, eds. Rohonezy, J., Parella, T. (BRUKER Analytik GMBH 1998-2000).

[53] Levitt, M. (2001) *Spin Dynamics* (John Wiley&Sons Ltd., Chichester).

[54] Cavanagh, J., Fairbrother, W. J., Palmer III, A. G. & Skelton, N. J. (1996) *Protein NMR Spectroscopy: Principles and Practice* (Academic Press, San Diego).

[55] Fernandez, D., Wider, D. (2003) *Current Opinion in Structural Biology*, 13:570–580.

[56] Wider, G., Wutrich, K. (1999) *Current Opinion in Structural Biology*, 9:594–601.

[57] Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).

[58] Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995) *J. Biomol. NMR* 6, 277-93.

[59] Johnson, B. A. & Blevins, R. A. (1994) *J. Biomol. NMR* 4, 603-614.

[60] Lepre CA, Moore JM. (1998) *J Biomol NMR*. 12(4):493-9.

[61] Kapust, R. B. & Waugh, D. S. (1999) *Protein Sci* **8,** 1668-1674.

[62] Cherry S., Tropea J. E., Waugh D. S.
http://mcl1.ncifcrf.gov/waugh_tech/protocols/pur_histev.pdf

[63] Saxena, P. & Walker, J. R. (1992) *J. Bacteriol.* **174,** 1956-1964.

[64] Kapust RB, Tozser J, Fox JD, Anderson DE, Cherry S, Copeland TD, Waugh DS. (2001) *Protein Eng*. (12):993-1000.

[65] Jelesarov I, Bosshard HR. (1999) *J Mol Recognit*. 12(1):3-18.

[66] Wishart, D. S., Sykes, B. D. & Richards, F. M. (1992) *Biochemistry* 31, 1647–1651.

[67] Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* 22, 4673-80.

[68] Risler, J. L., Delorme, M. O., Delacroix, H. & Henaut, A. (1988) *J. Mol. Biol.* 204, 1019-29.

[69] Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999) *Bioinformatics* 15, 305-8.